DOCUMENT RESUME

ED 420 663 TM 026 934

| | |
|---|---|
| AUTHOR | Glass, Gene V., Ed. |
| TITLE | Education Policy Analysis Archives, 1996. |
| REPORT NO | ISSN-1068-2341 |
| PUB DATE | 1997-04-00 |
| NOTE | 349p.; "Education Policy Analysis Archives" is an electronic-only journal (see http://olam.ed.asu.edu/epaa) covered on an article-by-article basis in "Current Index to Journals in Education" (CIJE). For the nineteen articles in this volume, see TM 520 174-192 (in CIJE). The numbered papers skip #18, numbering the last two papers #19 and #20. |
| PUB TYPE | Collected Works - Serials (022) |
| JOURNAL CIT | Educational Policy Analysis Archives; |
| EDRS PRICE | MF01/PC14 Plus Postage. |
| DESCRIPTORS | *Academic Achievement; Computer Uses in Education; Dropouts; Educational Assessment; Educational Change; *Educational Policy; Educational Trends; Electronic Journals; *Elementary Secondary Education; Foreign Countries; Graduate Surveys; Health Education; *Higher Education; Inclusive Schools; Mainstreaming; Policy Formation; *Professional Development; Psychometrics; Scholarly Journals; *Standards |
| IDENTIFIERS | England |

ABSTRACT

This document consists of downloaded copies of the nineteen (19) papers published in the Electronic Journal "Education Policy Analysis Archives" during 1996: (1) "The Achievement Crisis Is Real: A Review of 'The Manufactured Crisis'" (Lawrence C. Stedman); (2) "Staff Development Policy: Fuzzy Choices in an Imperfect Market" (Robert T. Stout); (3) "Making Molehills out of Molehills: Reply to Lawrence Stedman's Review of 'The Manufactured Crisis'" (David C. Berliner and Bruce J. Biddle); (4) "Standard Errors in Educational Assessment: A Policy Analysis Perspective" (Greg Camilli); (5) "Opening up Jewish Education to Inspection: The Impact of the OFSTED Inspection System in England" (Judy Keiner); (6) "The 1976 Illini: Sweet Memories of Alma Mater" (Diya Dutt); (7) "Respecting the Evidence: The Achievement Crisis Remains Real" (Lawrence C. Stedman); (8) "Developmentalism: An Obscure but Pervasive Restriction on Educational Improvement" (J. E. Stone); (9) "Markets versus Monopolies in Education: The Historical Evidence" (Andrew Coulson); (10) "Being Popular about National Standards: A Review of 'National Standards in American Education: A Citizen's Guide'" (Michael W. Apple); (11) "National Education 'Goals 2000': Some Disastrous Unintended Consequences" (Robert H. Seidman); (12) "Public School Reform: Potential Lessons from the Truly Departed" (J. Dan Marshall and James P. Valle); (13) "Implementing AIDS Education: Policies and Practices" (Grace C. Huerta); (14) "Actual Schools, Possible Practices: New Directions in Professional Development" (Rebecca Novick); (15) "A Review of Dorn's 'Creating the Dropout'" (Aimee Howeley); (16) "A Review of 'Computers as Tutors: Solving the Crisis in Education'" (Greg Sherman); (17) "What Does the Psychometrician's Classroom Look Like? Reframing Assessment Concepts in the Context of Learning" (Catherine S. Taylor and Susan Bobbitt Nolen); (18) "Inclusive Education in the United States: Beliefs and Practices among Middle School Principals and Teachers" (C. Kenneth Tanner, Deborah Jan Vaughn

Linscott, and Susan Allan Galis); and (19) "'The Bell Curve': Corrected for Skew" (Haggai Kupermintz). (SLD)

********************************************************************************
*          Reproductions supplied by EDRS are the best that can be made        *
*                         from the original document.                          *
********************************************************************************

ED 420 663

VOLUME 4 - April 1996

ISSN 1068-2341

# Education Policy Analysis Archives

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Gene V. Glass

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM026934

**editor:**
**Gene V Glass**
**Arizona State University**

# Education Policy Analysis Archives
## Volume 4

**Total Pages Per Article**

1 The Achievement Crisis Is Real: a Review of the Manufactured Crisis . . . 11
2 Staff Development Policy: Fuzzy Choices in an Imperfect Market . . . 15
3 Making Molehills out of Molehills: Reply to Lawrence Stedman's Review of the Manufactured Crisis . . . 13
4 Standard Errors in Educational Assessment: a Policy Analysis Perspective . . . 17
5 Opening up Jewish Education to Inspection: the Impact of the Ofsted Inspection System in England . . . 21
6 The 1976 Illin: Sweet Memories of Alma Mater . . . 15
7 Respecting the Evidence: the Achievement Crisis Remains Real . . . 29
8 Developmentalism: an Obscure but Pervasive Restriction on Educational Improvement . . . 29
9 Markets Versus Monopolies in Education: the Historical Evidence . . . 25
10 Being Popular about National Standards: a Review of National Standards In American Education: a Citizen's Guide . . . 6
11 National Education "Goals 2000": Some Disatrous Unintended Consequences . . . 35
12 Public School Reform: Potential Lessons from the Truly Departed . . . 12
13 Implementing AIDS Education: Policies and Practices . . . 16
14 Actual Schools, Possible Practices: New Directions in Professional Development . . . 15
15 A Review of Dorn's Creating the Droupout . . . 5
16 A Reveiw of Computers as Tutors: Solving the Crisis in Education . . . 5
17 What Does the Psychometrician's Classroom Look Like? Reframing Assessment Concepts in the Context of Learning . . . 35
18 Inclusive Education In the United States: Beliefs and Practices among Middle School Principals and Teachers . . . 30
19 The Bell Curve: Corrected for Skew . . . 11

**Education Policy Analysis Archives**

**Volume 4 Number 1** **January 23, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS
ARCHIVES.Permission is hereby granted to copy any
article provided that EDU POLICY ANALYSIS
ARCHIVES is credited and copies are not sold.

# The Achievement Crisis is Real:
# A Review of *The Manufactured Crisis*

**Lawrence C. Stedman**
**State University of New York-Binghamton**

*stedman@bingsuns.cc.binghamton.edu*

**Abstract:** In a provocative new book, *The Manufactured Crisis,* David Berliner and Bruce Biddle make four sweeping claims about U.S. achievement:

- there never was a test score decline,
- today's students are "out-achieving their parents substantially" (p. 33),
- U.S. students "stack up very well" in international assessments (p. 63), and
- the general education crisis is a right-wing fabrication.

As a progressive, I'm sympathetic to their concerns, but as a scholar who specializes in this material, I find their analysis deeply flawed and misleading. They mischaracterize the test score decline data, mishandle the international findings, and fail to acknowledge students' continuing low levels of academic achievement.

### The Decline

Although Berliner and Biddle are generally right that achievement has been stable, they ignored important contradictory evidence and the 1970s decline. They claimed "only 'one' test, the SAT" ever suggested a decline (p. 35). This is remarkable. High school students' NAEP civics scores, for example, dropped substantially between 1969 and 1976 and have been slipping ever since. Their science scores also fell during the 1970s and have only partly rebounded. Several commercial tests, such as CTBS and STEP, showed declines in the 1970s. In the late 1980s, senior high school reading scores declined on the MAT while reading and math scores fell in many grades on the SRA (Linn, Graue, & Sanders, 1990). In the late 1980s, younger students' NAEP reading and writing performance slipped. (For details, see Stedman & Kaestle, 1991; Stedman, 1993.)

They attributed the SAT decline to demographic changes in test takers, yet never reviewed the evidence which shows this explains much, but not all, of the decline. They used "average" SAT scores to claim minority gains, but this masked minority verbal declines in the late 1970s and late 1980s (Stedman, 1994b). Mexican-American, Puerto-Rican, and Asian American verbal scores were about the same in the early 1990s as they were in 1976.

Berliner and Biddle made sweeping claims about recent gains on commercial tests. Their handling of the Linn, Graue and Sanders study demonstrates how selective they are with evidence. Their graph omitted Linn, Graue and Sanders' SRA data which showed declines in many grades. They only graphed the elementary school data, which hid the less impressive high school scores, some of which were declining or stagnating. They never mentioned that Linn, Graue and Sanders pondered, "But the more important question is: Has student 'achievement' improved in recent years?" and concluded that the answer was "equivocal" (Linn, Graue & Sanders, 1990, p. 13). Linn, Graue and Sanders determined that recent gains were partly caused by districts' repeated use of the same tests rather than by genuine improvement. The 1980s back-to- basics movements also artificially raised scores by frequent testing and skill-drill approaches (Stedman & Kaestle, 1991).

Finally, Berliner and Biddle claimed "virtually all" commercial tests would "show that today's students are out- achieving their parents substantially" (p. 33), yet never presented any evidence to support their claim. They ignored the many reviews of historical trends on equating studies which refute their claim (Stedman & Kaestle, 1987). The best that can be concluded is that this generation of students "generally" performs about the same as earlier ones, but the patterns are complicated and there is contradictory evidence.

Given changing school populations and societal conditions, generally stable scores are still a remarkable accomplishment for U.S. schools. This is an important message that the public needs to hear. Nevertheless, the reality is more complicated than they suggested. Although school critics often exaggerated the extent and ramifications of the declines, many did occur (Stedman and Kaestle, 1991). Berliner and Biddle should have admitted that, on several indicators, our students are not performing as well as they once did.

### International Assessments

U.S. performance in the international arena is not as dismal as school critics have asserted, but it certainly is not as glowing as Berliner and Biddle claim. Our students have done well in reading and elementary school science, middling to poor in geography and secondary school science, and last or near-last in mathematics (Stedman, 1994b). Berliner and Biddle offered several arguments to try to explain the weak U.S. performance but, in doing so, they tacitly acknowledged that our international performance often has been poor.

### Opportunity-to-Learn

Berliner and Biddle's opportunity-to-learn argument is a red herring. International researchers pioneered the use of OTL measures and it is already factored into many results. ETS's 1988 international math and science findings, for example, came only from schools in which "more than 75 percent of the students had already had an opportunity to learn the content" (Lapointe et al., 1989, p. 33). Even so, the U.S. did poorly whether judged by rankings, proficiency levels, or percentage correct.

Berliner and Biddle claimed that our students are at a disadvantage because we generally delay algebra until 9th or 10th grade. But U.S. students have done poorly in most math areas, not just algebra. In 1988, for example, our 13-year-olds ranked last in arithmetic and measurement and next-to-last in geometry, data organization, and problem solving (NCES, 1991, p. 395). They also had poor results in 1991 in these areas (NCES, 1992, p. 21).

U.S. and Japanese curricula were also more comparable than claimed. In the Second International Mathematics Study, content coverage was similar in arithmetic, geometry, and statistics, yet U.S. students still scored lower (Stedman, 1994a). In a telling analysis, Baker (1993) found that when one considers "only" the test items that U.S. 8th graders were taught during the year, they averaged only 40% correct.

### Westbury Study

The Westbury study was at the heart of their curricular claims, but their handling of it revealed they care more about their argument than the evidence. First, the study has limited implications because it used data that were over a decade old, dealt only with one subject--math, and involved a better-than- usual U.S. 8th grade performance. Second, they did not even report Westbury's comparisons properly. They took his scores for our most advanced 8th grade math classes--the top 25% comprising algebra and pre-algebra--and compared them to the "average" Japanese class! No wonder our algebra classes looked good in their comparison.

What Westbury actually did was compare our most advanced 8th grade math classes to the top fifth of Japanese students. Although this was a fairer approach, it still did not "isolate" the effects of the curriculum, but confounded them with selection effects. U.S. students who study algebra in 8th grade are a select group of 14%, differing from other U.S. students in college expectations, math interest, parental support, social class, and academic ethic. Consequently, one cannot tell how much of their performance reflects their algebra curriculum and how much their background advantages. (Using this comparison directly violated their own research precept--the Principle of Control, p. 159.)

What did Westbury actually find? Our select students did not do that well. Our pre-algebra classes scored only 56% correct and lagged well behind, by a substantial two standard deviations (Westbury, 1992, p. 21). Our algebra classes scored comparably to the Japanese classes, but this was hardly surprising. They were an elite group of only 14% of our classes compared to a less select 20% of the Japanese students. They were judged only on the algebra portion of the test, yet they had spent more of their time on algebra (formulas and equations), 61% to 26%, and had covered more of the test problems, 88% to 82%, than the Japanese students (Westbury, 1992, p. 20, p. 21). (So much for claims that curricula were equated!). In two other test areas, geometry and measurement, they even scored below the "average" Japanese class (Stedman, 1994a). Finally, our 8th graders were older and had been in school longer--the Japanese students were only 7th graders!

Berliner and Biddle ignored Westbury's analysis of U.S. calculus classes, yet this tested the overall quality of our best math programs given to our best students. Our calculus classes fared poorly, however, substantially trailing the "average" Japanese class in every tested area (Stedman, 1994a). Given all this, it was misleading for them to claim that "U.S. teachers and schools are [not] deficient compared with those in Japan" (p. 56) and to conclude that "Many, perhaps most, of the studies' results were generated by differences in curricula" (p. 63).

### Variability Argument

Berliner and Biddle tried to explain away poor U.S. international performance by claiming our achievement is "a 'lot' more variable" (p. 58) than other countries, but offered no evidence. In fact, the 1991 IAEP math and science studies showed our variability was similar to that of other nations and less than that of Taiwan and Korea, the leading performers (cf. 10th & 90th percentiles, NCES, 1993b, p. 56; NCES, 1993a, p. 415).

### States-to-Nations Comparison

They never mentioned that the states-to-nations comparison they cited was designated "experimental" and technically problematic (see caution, NCES, 1993b, pp. 54, 94). The international scores were projections from a U.S. sample that took both the NAEP and IAEP tests. No international student ever took the NAEP test and it is unclear that the IAEP-NAEP relationship would be the same for students in other countries. Our states had two important advantages. Our students were older--over half were 14-15 years old whereas the international students were 13-year- olds. Our states' scores came from the 1992 NAEP assessment and were higher than what was projected for the U.S. (cf. NCES, 1993c, p. 83; NCES, 1993b, p. 56).

Finding that a few select, typically high-scoring mid-Western states did well in the comparison is not surprising. What is staggering is that our best state scores were only the "average" level in Taiwan and Korea! Berliner and Biddle did not report that the same comparison showed that the typical U.S. student was two years behind the average Taiwanese student and scored only around Taiwan's and Korea's 25th percentile (NCES, 1993b, pp. 54, 56). It also showed that only 13-16% of U.S. students reached the proficient level, while 35-43% of Taiwanese and Korean students did (Pashley & Phillips, 1993).

### Social Inequality Argument

Although racism and social inequality have taken a severe toll on many of our students' academic development, this does not explain the poor general performance of U.S. students. The math deficit, for example, is not simply a minority student problem. In 1992, only 30% of "white" U.S. 8th graders demonstrated proficiency in the NAEP math assessment; over a quarter did not even make the basic level (NCES, 1993c, pp. 101-102). Nor are our problems due to low-achievers. Even our top half have not kept pace internationally in math and science (Stedman, 1994a).

Although U.S. students do not generally fail in international comparisons, it is misleading for Berliner and Biddle to claim that "they stack up very well" (p. 63).

### Low Achievement

The book's central problem is that Berliner and Biddle tell only part of the story. Although achievement trends, for the most part, have been stable, academic and general knowledge have been at low levels for decades (Stedman, 1993).

In math, NAEP analysts recently concluded that "less than half (of high school seniors) appeared to have a firm grasp of seventh-grade content" (Mullis et al., 1991, p. 80). They have trouble even with simple problems involving fractions, decimals, and percents.

Few high school students have done well on NAEP writing tests. Only about a third wrote adequate papers and only a small percentage could write "elaborated" papers. The one bright spot is their competence in basic grammar and punctuation.

Our functional illiteracy rate remains around 20-30%--meaning that millions of adults have trouble with common day-to-day reading tasks (Stedman & Kaestle, 1987; Kirsch, 1993).

Students lack basic knowledge in history and literature. In the late 1980s, substantial majorities of our 17-year-olds did not recognize that Upton Sinclair was a muckraker, the Scopes trial dealt with evolution, Jim Crow laws segregated blacks, or the time period of the Civil War. A majority did not recognize classics by Shakespeare, Chaucer, Conrad, and Whitman and were unfamiliar with major women and African-American writers. These were straight- forward multiple-choice questions deliberately designed without the usual distractors.

Geographical knowledge also has often been poor. In 1988, Gallup repeated a survey given to adults in 1947 and concluded that "Americans' geographic literacy has gotten worse in the last forty years." They found that, "From outline maps, the average American can identify only four of twelve European countries, less than three of eight South American countries, and less than six of ten U.S. states" (Gallup Organization, 1989, p. 162).

### Rejoinders

Instead of reviewing and acknowledging this evidence, Berliner and Biddle offer several rejoinders why such findings don't matter. They suggest that the standards for knowledge are unrealistic and are those of classicists, historians, and test designers. Most people, however, would expect high school seniors to be competent in 7th grade math, literacy, and basic social studies information.

### Breadth of Experience

They argue that U.S. students are focused on a breadth of experience, but this does not excuse our low achievement. Certainly academic achievement is one of our goals and should be one of our strengths. Nor is it clear that U.S. students have a monopoly on breadth or richness of experience. Portraits of Japanese elementary schools clearly show that students are not academic automatons, but are engaged in rich curricular and extra- curricular activities--calligraphy, sewing, hands-on math and science activities, group problem-solving, electronics, dance, musical training, play, reading, physical exercise, cooperative learning, school jobs, etc. (Stevenson & Stigler, 1992).

### Scaling Problems

They rightly argue NAEP scales are flawed, but this does not explain students' poor performance or limited knowledge. Contrary to their assertions, it doesn't require tough questions to generate scale scores or discriminate among U.S. students. The problems at the highest NAEP levels are actually fairly easy. The 300 level in math, for example, includes simple decimal problems and level 350 has "routine problems involving fractions and percents." This is junior high general math, yet 17-year-olds have trouble with it! In history, many 350 level problems required nothing more than simple recognition of basic facts. (For a more detailed look at NAEP findings as well as its scaling problems, see Stedman, 1993.)

Many findings of low performance do not come from traditionally scaled tests. The writing results involve authentic holistic evaluations and thus avoid the scaling problems. The functional illiteracy estimate came from tests of many different designs and was derived through a systematic analysis of individual items not scaled results. The true rate might even be higher because some tests used items that were easier than their real-life counterparts and did not test dropouts, the homeless, prisoners, or non-English speakers (Stedman and Kaestle, 1987). Low levels of civic literacy and general knowledge were revealed in national surveys as well as standardized tests.

### Details of Low Achievement

Careful reviews of individual items and sets of items have avoided many scaling problems and still indicate students struggle with basic material (Carpenter et al, 1988, 1982). Math educators found that students "exhibit serious gaps in their knowledge" and often learn "concepts and skills at a superficial level" They concluded that "students' achievement at all age levels shows major deficiencies" (Carpenter et al, 1988, pp. 40-41). In 1990, for example, only around half the 17-year-olds could convert a decimal to a fraction, find a number given a percent, estimate a square root, and use the properties of triangles (Mullis et al, 1991, pp. 302-309). 34% could not even find the area of a rectangle, given a diagram and the length of two sides (Mullis et al, 1991, p. 306).

Although students' geographical knowledge is better than many have asserted, there still are serious problems (Stedman, 1993). 15-40% of high school students had trouble with basic geographical material. Most could not interpret a graph showing birth and death rates. Given the Vietnam War, it is unsettling that 63% of our high school seniors could not locate Southeast Asia on a world map. 64% did not know Saudi Arabia's location, although this was before the Persian Gulf War. Half could not answer such simple questions as the following:

> The construction of the Panama Canal shortened the sailing time between New York

and [London, Port-au-Prince, Rio de Janeiro, San Francisco]

Functional literacy tests have produced some disturbing findings. Twenty percent of the population, for example, had trouble reading and understanding dosage information on medicine bottles. Similar percentages had problems with a housing inspection notice, basic coupons, and price per unit weight. About a third failed at figuring out train schedules, how much change should come from a purchase, and which subjects had improved on a report card.

Student achievement may be even worse than these findings suggest. The NAEP data do not include dropouts who presumably would score lower. To reach a given NAEP level, students only have to answer correctly 65-80% of its problems. The burden on students is light. Compared to the SATs and achievement tests, which can be half-day or all-day affairs, the NAEP tests are short, only 45 minutes. The tests are predominantly multiple choice, recognition-based rather than open-ended, recall which make them easier for students to do well on.

**Real-World Relevance**

Berliner and Biddle argue that findings about low achievement are irrelevant because the tests did not measure real-world problem solving. This is a curious position given that their claims about stable achievement trends came from these same tests! There are several problems with their argument. First, many tests that showed low achievement did measure the knowledge and skills needed in the real world. The functional literacy tests, for example, used real-world tasks with real-world materials. Math tests have involved calculators, graphing, and open-ended items. NAEP reading tests have used poetry, newspaper articles, and passages from real literature.

Second, in-school and out-of-school tasks, although different in many ways, still involve related abilities. Standardized tests give some indication of real-world problem solving ability. One indication of this is the marked correlation between scores on traditional tests and those on authentic assessment measures (Wang, Haertel, and Walberg, 1993, p. 371).

Third, "real-world problem-solving" is not our only educational goal. General knowledge, some of which can be measured successfully via multiple-choice testing, is an important goal in itself. We want informed and knowledgeable citizens. Historical knowledge can play a central role in understanding public policy debates.

Consider the on-going and highly-charged debates over immigration policy and affirmative action. How can we expect students and young adults to make informed appraisals of the arguments when they are ignorant about the history of race relations in this country? NAEP testing in the 1980s showed that the vast majority of high school students did not know what Jim Crow laws were, what the 3/5ths Compromise was, or what the Emancipation Proclamation actually did. Substantial percentages did not know what Plessy v. Ferguson or Brown v. Board of Education were about. They lacked basic information pertaining to the Civil War, one of our nation's epochal events and a key force in shaping race relations. Sizable majorities were unfamiliar with the Missouri Compromise, nullification, the Dred Scott decision, the dates of Civil War, and the dates of Lincoln's term. Such ignorance is not an artifact of an obscure psychometric scaling procedure. Knowledge does matter.

Fourth, there likely will be little comfort in results from more authentic, real-world testing. NAEP is increasingly using performance assessment--the new science test, for example, includes drawing tasks, writing, and open-ended questions. The new reading assessment has longer passages and is 40% open-ended. Students, however, often do more poorly on the open-ended versions of test items. When their understanding of a subject is probed, surprising gaps and confusions often appear (Bridgeman, 1992; Martinez, 1991; NAEP, 1983, p. 32; Rogers & Stevenson, 1988). Future assessments are likely to produce even more disturbing news about low achievement than we have now.

Finally, Berliner and Biddle argue that school critics focused more on the imagined economic consequences of low achievement than on the actual achievement evidence. I agree. Soon after the "Nation at Risk" report appeared, I argued that it made too much of a high-skilled, hi-tech future economy as a rationale for reforming education (Stedman and Smith, 1983). But the actual evidence is troubling and Berliner and Biddle did not examine it. The low levels of achievement are unimpressive results for 12 years of schooling. The tests do measure much of what is being taught in our schools and show we are not succeeding in our efforts. A complex, democratic society needs a well-read and knowledgeable citizenry and yet the evidence shows we are not accomplishing this.

### Teaching Methods and Student Work Habits

Our achievement problems are deep-seated, widespread, and long-standing. But this is not the only reason for fundamental and far-reaching school reform. Teaching methods and student work habits also leave much to be desired (Stedman, 1993). Although there are a few bright spots, such as the frequent use of demonstrations in science classes, the portrait is troubling. NAEP analysts found math instruction

> "continues to be dominated by teacher explanations, chalkboard presentations, and reliance on textbooks and workbooks. More innovative forms of instruction--such as those involving small group activities, laboratory work, and special projects--remain disappointingly rare." (Dossey et al, 1988)

History and civics classes are dominated by textbooks, tests, quizzes, and short-answer questions. It is unusual to find students working in groups or writing long papers. Writing instruction in the schools is also limited and is focused on mechanics. Only about a fourth of 8th graders report that their teachers spend more than an hour a week on writing.

Interest in science has not been sparked. In 1986, fewer than a fourth of 11th graders reported working on science-related hobbies or talking with friends about science. Only about a third reported going to a science museum or trying to fix something electrical or mechanical.

Students do little schoolwork. The data on homework and TV watching are revealing. In 1990, only about a third of our 17- year-olds reported spending over an hour a day on homework, whereas half reported watching 3 or more hours of TV daily! Reading has been shortchanged. In 1986, over half the 11th graders reported reading on their own less than once a week; about a fifth reported they never did!

One cannot look over this information without a sense that our schools are not what they should be. Over the past decade, thought-provoking ethnographies and school profiles by Boyer, Fine, Goodlad, Oakes, Sizer, and others have portrayed a school system in crisis. What we're seeing, particularly at the high school level, is that students are often disengaged, teachers' work is often factory-like, and intellectual life is often poor. These accounts were hardly the products of right-wing ideologies (cf. Berliner & Biddle, Chapter 4, pp. 140-141).

Reformers have been busy. They know that the schools are not better than ever, but rather, more than ever, they need to be different than they are. Teachers and other educators who are intimately involved in the life of schools recognize there is a serious problem. There are major reform efforts affecting every major aspect of education: curriculum, evaluation, funding, governance, pedagogy, and school organization. Local educators are not mere pawns in a conservative political chess game, but have been responding actively to real needs and problems.

### The Scope of Reform

Fixing the schools is a crucial part of solving our long- standing academic problems. But we also need to create a society that values scholarship and learning over commercialism and entertainment. This will require a major political and economic transformation.

Educators must challenge the vested interests that are more interested in profits than the welfare of communities and civil society. We must fight the economic displacements that disrupt families, produce violence, and undermine students' development. We must take on the media conglomerates that are focused more on selling products than nurturing our cultural and intellectual life. We must change a system that values the bombastic broadsides of radio talk show hosts and political candidates over reasoned and civil discourse.

To succeed in our most troubled communities, we will need to overhaul school financing systems and break down powerful, entrenched bureaucracies. But school reform is no substitute for job creation, income redistribution, and political empowerment. We must make our educational efforts part of a broader social and political agenda, one that promotes full employment, community revitalization, and civic participation.

### Conclusion

In the 1980s, school critics often exaggerated the size and extent of the test score decline. In spite of enormous changes in society and school populations, U.S. achievement has been remarkably stable for many decades. But it remains inadequate and at low levels. Ignoring this evidence or arguing it is a right- wing fabrication hampers much needed school reform. The crisis is real, what is actually being manufactured here is a new mythology about U.S. student achievement.

Interested readers can find an in-depth--and balanced-- treatment of the achievement evidence in several of my recent articles. I examine the NAEP data in the November 1993 Phi Delta Kappan, the Sandia Report and the SAT data in the January-February 1994 Journal of Educational Research, and the international data in the October 1994 Educational Researcher.

## References

Baker, D. P. (1993). Compared to Japan, the U.S. is a low achiever Really. Educational Researcher, 22(3), 18-20.

Bridgeman, B. (1922, Fall). A comparison of quantitative questions in open-ended and multiple-choice formats. Journal of Educational Measurement, 269+.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Jr., Lindquist, M. M., & Reys, R. E. (1982). Student performance in algebra: Results from the National Assessment. School Science and Mathematics (October), 514-531.

Carpenter, T. P., Lindquist, M. M., Brown, C. A., Kouba, V. L., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP assessment of mathematics. Arithmetic Teacher (December), 38-41.

Dossey, J. et al. (1988). The Mathematics Report Card. Princeton, New Jersey: Educational Testing Service.

Gallup Organization (1989). The Gallup Poll public opinion 1988. Wilmington, Delaware: Scholarly Resources.

Kirsch, I. S. (1993). Adult literacy in America. Washington, D.C.: National Center for Education Statistics.

Lapointe, A., Mead, N., & Phillips, G. (1989). A world of differences. Princeton, New Jersey: Educational Testing Service.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district results to national norms: The validity of claims that "everyone is above average". Educational Measurement: Issues and Practice (Fall), 5-14.

Martinez, M. (1991, Summer). A Comparison of Multiple-Choice and Constructed Figural Response Items. Journal of Educational Measurement, 131-45.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991). The Sate of Mathematics Achievement. Washington, D.C.: U.S. Department of Education.

NAEP (1983). The Third National Mathematics Assessment. Report No. 13-MA-01. Denver, Colorado: Education Commission of the States.

National Center for Education Statistics (1993a). Digest of education statistics 1993. Washington, D.C.: U.S. Department of Education.

National Center for Education Statistics (1993b). Education in States and Nations. Washington, D.C.: U.S. Department of Education.

National Center for Education Statistics (1993c). NAEP 1992 Mathematics Report Card for the Nation and the States. Washington, D.C.: U.S. Department of Education.

Pashley, P. J. & Phillips, G. W. (1993). Toward world-class standards: A research study linking international and national assessments. Princeton, New Jersey: Educational Testing Service.

Rogers, V. & Stevenson, C. (1988, February). How Do We Know What Kids Are Learning in School? Educational Leadership, 68-75.

Stedman, L. C. (1993). The condition of education: Why school reformers are on the right track. Phi Delta Kappan, 75 (3), 215-225.

Stedman, L. C. (1994a). Incomplete explanations: The case of U.S. performance in the international assessments of education. Educational Researcher, 23(7), 24-32.

Stedman, L. C. (1994b). The Sandia Report and U.S. achievement: An assessment. Journal of Educational Research (January- February), 133-146.

Stedman, L. C. (1995). The new mythology about the status of U.S. schools. Educational Leadership, 52(5), 80-85.

Stedman, L. C. & Kaestle, C. F. (1987). Literacy and reading performance in the United States, from 1880 to the present. Reading Research Quarterly (Winter), 8-46.

Stedman, L. C. & Kaestle, C. F. (1991). The great test score decline: A closer look. In C. F. Kaestle, H. Damon-Moore, L. C. Stedman, K. Tinsley, & W. V. Trollinger (Eds.), Literacy in the United States (Chapter 4). New Haven: Yale University Press.

Stedman, L. C. & Smith, M. S. (1983). Recent reform proposals for American education. Contemporary Education Review, 2(2), 85- 104.

Stevenson, H. & Stigler, J. (1992). The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education. New York: Simon and Schuster.

Wang, M, Haertel, G., & Walberg, H. (1993). Toward a knowledge base: Why, how, for whom? Review of Educational Research, 63(3), 365-376.

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? Educational Researcher, 21(5), 18-24.

## About the Author

### Lawrence C. Stedman

stedman@binghamton.edu

Lawrence C. Stedman is Associate Professor of Education at the State University of New York at Binghamton. His Ph.D. is from the University of Wisconsin at Madison in Educational Policy Studies with a minor in Sociology. He has worked as a school district policy analyst, secondary school teacher, VISTA volunteer, and educational researcher. He has a keen interest in equal opportunity and school reform. His dissertation and early articles centered on effective schools research and the reform reports of the early 1980s. He has helped evaluate ESL, minority achievement, merit pay, and dropout intervention programs.

More recently, his research has focused on the general condition of education and its implications for policy-making. He has written articles on the test score decline, literacy trends, the international assessments, and the Sandia Report. He is currently investigating historical trends in students' and adults' general knowledge. It is the outgrowth of a book he helped author with Carl Kaestle and others on the history of the U.S. reading public (*Literacy in the United States: Readers and Reading Since 1880*, Yale University Press, 1991). This new research has been funded by a SUNY Faculty Research Grant and Fellowship and by a National Academy of Education Spencer Foundation post-doctoal fellowship.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Mark E. Fetler
*fetlerctc.aol.com*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 2** **February 17, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

# Staff Development Policy: Fuzzy Choices in an Imperfect Market

**Robert T. Stout**
**Arizona State University**

*stout@asu.edu*

**Abstract:** It is argued here that staff development in the public elementary and secondary schools of the United States is misguided in both policy and practice. In its current form it represents an imperfect consumer market in which "proof of purchase" substitutes for investment in either school improvement or individual development. A policy model based on investment in school improvement is shown, in which different assumptions about how to improve schools are linked to different alternatives for the design and implementation of staff development. These are argued to be based on an investment rather than consumption model.

Public policy about staff development for teachers is confused by both lack of clear purpose and by unsatisfactory decision criteria. Lanier and Little (1986) concluded that "staff development has not generally been the product of coherent policy, nor has it been systematically integrated with institutional priorities for curriculum and instructional improvement" (p. 562). Consequently, policy makers have little opportunity to assess either costs or benefits of what is a large public investment. Nonetheless they continue to view staff development-sometimes called continuing education, in-service training, or professional development-as a basic tool for changing teacher behaviors, and therefore schools. The view may be misplaced or wrong-headed but it prevails.

But fundamental policy choices exist. If they were made apparent they might lead to modifications in public policy decisions about investment in staff development. Mitchell (1986), for example, argued if school leaders believe that improving individual skill or motivation among teachers is more likely to improve performance than close supervision, staff development may be a good school improvement strategy. Further, when leadership assumes that the average level of teacher performance is shaped more by cultural beliefs and subjective feelings than by objective work conditions (like class size or textbook quality) staff development may become a favored tactic.

Although current staff development policy is muddled at best, and out of control at worst, policy makers continue to assert the value of the enterprise. Almost all states in the United States require some form of continuing education for teachers. The major national reports on teacher reform, Tomorrow's Teachers (Holmes Group, 1986) and A Nation Prepared: Teachers for the 21st Century (Carnegie, 1986), emphasize the need for teachers to continue to learn. Virtually every school district in the country provides some form of staff development for teachers. Salary schedules, merit pay schemes, and career ladders throughout the country reward teachers for participation in staff development.

## TEACHER MOTIVATION TO PARTICIPATE

But why would teachers bother to participate? At least four motives underlie teacher decisions to do so. One is salary enhancement. Participation pays off. Automatic salary raises often accrue quickly, and almost always eventually. Eligibility to compete for merit pay or to climb a career ladder are often tied to "demonstrated commitment to personal and professional development" (read participation in staff development). Another motive is certificate maintenance. State policy makers assume, whether rightly or wrongly, that periodic retooling is desirable and that continuing in the occupation should be dependent on it. A third motive is career mobility. Teachers take courses and degrees and participate in workshops to build resumes. Having done so, they attempt to leave education for other occupations or to pursue other careers within education, administration being the notable example.

None of these three motives, in itself, necessarily leads to better performance. Sometimes participation will do so, but nothing exists in the system to ensure, or perhaps even encourage, it. If a teacher's skills improve, and if the enhanced skill can be shown to result in higher levels of student performance, or any other measure of school output, then policy assumptions have been satisfied. But on the face of it, the evidence is missing that staff development, as currently arranged, can produce these links.

Teachers talk about the fourth motive, but in vague terms. Almost always the language is of gaining new skills/knowledge to enhance classroom performance. The motive is both noble, and appropriate, from a public policy perspective. The problem, as will be argued, is that the chances of policy-appropriate motive connecting to available, timely, and intellectually honest sources are little more than accidental.

## A SHORT EXCURSION INTO COST

Although computing the costs of all of this activity is beyond the scope of this paper, a review of the sources of cost may be instructive. The costs are both direct and indirect. Direct costs accrue as a result of direct payment to service providers. State-level costs occur when the state education agency (1) provides direct staff development assistance, (2) funds project-related efforts by local education agencies or by private providers, or (3) processes advanced certification requests. Every state in the country employs department of education specialists who provide direct service to school districts, and each state has officials who process information about teacher progress toward this or that advanced certificate. County offices of education (or some other intermediate unit) often provide direct services which parallel or supplement those of the state education agency.

Local education agencies also incur direct costs. Smaller costs manifest in one or two days a year set aside as "in-service days." A higher level of cost in incurred when a local education agency decides to stress a particular theme. In this scenario, a private consultant (or district employee) provides a series of workshops and other training experiences. Perhaps the greatest direct expenditure of local funds attaches to general school improvement efforts, in which teachers are instructed in techniques, planning, curriculum development, and a host of other topics associated with a general model of school change.

Additional direct costs are incurred when substitute teachers must be hired to replace teachers who are attending staff development activities. Although much staff development takes place before and after regular school hours, sufficient activity occurs during the work day that this cost is noticeable. A final direct cost is that borne by teachers. As they advance in certification or obtain advanced degrees, teachers incur out- of-pocket expenses for college and university course work. Although the costs of such courses are recovered manyfold (given contemporary salary schedules), teachers nonetheless incur them.

The most expensive indirect cost of staff development rests with typical school district compensation systems. Under these, teachers get automatic pay raises for completing courses and workshops offered by whatever agencies are recognized by the school district. A related indirect cost is that public subsidies are provided for many of the providers. Public universities and colleges may enjoy subsidies of 30 percent or more of the true direct costs of instruction. Presumably, independent colleges and universities show profit, but even they receive public subsidy through the use of school-district facilities or by hiring full-time employees of local school districts and paying them modest wages.

Although beyond the scope of this work to establish, the total cost of all of this is not trivial. Little (1988) estimated that in 1986-87 staff development costs (direct and indirect) in California were about $368 million, or about $1700 per certified staff member. This figure is consistent with that reported by Miller, Lord and Dorney (1994), who estimated costs between $1700 and $3500 per teacher in four school districts.

## THE PUBLIC INTEREST

Although pubic interest in staff development is long-standing, shifts of focus and authority have been common, reflecting, perhaps, a continuing uncertainty over purpose, and discomfort about quality.

Two general policy goals have been associated with staff development in this century; general upgrading of teacher skills and preparing teachers to accomplish new tasks (Stout and Wigand 1982). The locus of policy interest has shifted from the states and state interests in insisting that teachers be college graduates, to the federal government, and back to an alliance between state and local policy makers. Federal interest was at its peak during the years from about 1956 to about 1975, during which time staff development was used as a mechanism to produce a general reformation of America's schools. Better curricula and better personnel were thought to be much needed and to be possible through federal intervention in training. Now state and local policy makers have received and responded to the mandate to recapture excellence and are using staff development in a host of ways. At the core of these efforts is a belief that staff development can produce school improvement.

Over the past sixty or so years, policy about staff development has not been guided by a single consistent purpose. Row and column salary schedules have been used to improve the teaching force in a general way. Some targeted efforts have been implemented in response to changes in federal policy directions, and periodic efforts have been made to link staff development to systematic school reform efforts. But, overall, these efforts have been without general direction and the coordination required to achieve some clear purpose.

## THE MARKET SYSTEM FOR STAFF DEVELOPMENT

The lack of policy focus in staff development is confounded by the nature of the market system through which it is provided. The multiple motives of teachers to participate have been described and the assertion made that only one of the four (the most difficult to track) has a clear potential link to improving school performance. In addition, the system of providing staff development is not unlike a giant academic bazaar. Colleges and universities compete fiercely for clients. In metropolitan areas of any size, tens of colleges and universities may be offering courses with the same title at the same time. Thousands of other providers crowd the marketplace as well. Local education agencies, county and state education agencies, private consultants, publishers and manufacturers of instructional materials, and purveyors of all sorts of answers to education concerns and problems set up their stalls and attempt to attract paying customers. This market is largely unregulated with respect to quality, though it is regulated in part with respect to form. The absence of quality controls is a result of both the absence of a clearly understood purpose and the motive systems that induce teachers to participate.

States have attempted to address the question of quality and return on state subsidy by regulating processes and procedures such as the number of required contact hours for courses or mandating that examinations be given in them. Some states have, from time to time, mandated content, particularly in response to a hot curricular issue. In other states, state agency employees have entered the marketplace as competitors. While these actions have encumbered and complicated the marketplace, they are misplaced because they are based on a misunderstanding of the operating market mechanism.

Staff development is a consumer market, albeit an imperfect one. In a true consumer market quality derives from consumer expectations of benefit and subsequent consumer choice. Bad products are driven out by consumer disinterest because the product is expected to produce utility for the purchaser. Products which do not are not purchased. In the staff development market, however, the inherent utility of a course or activity is irrelevant. The utility does not lie in the experience, but in evidence that the experience has been purchased. The consumer market analog is the "proof of purchase" which can be redeemed for a rebate or premium. In the case of staff development the "proof of purchase" is a transcript showing course completion, or a degree, or a certificate of attendance. The "proof of purchase" is traded for utility. Consequently, quality of the experience is easily sacrificed by participants for convenience or ease of access or free parking or a host of other considerations. Three of the motives to participate (salary enhancement, certificate maintenance, career mobility) are satisfied by showing sufficient numbers of proofs of purchase. At the point of "cashing in," proofs of purchase from one experience or course or institution are as good as those from any other.

The market then is a high volume, high cost, consumer driven one in which utility is disconnected to product and current regulatory attempts are misplaced because they do not affect the primary currency.

## APPARENT CONSEQUENCES OF THE CURRENT MODEL

Teachers in large numbers continue to participate in staff development, and providers have multiplied as creative people continue to develop new delivery systems. Yet research evidence continues to be elusive, with no demonstrated link between teacher performance and attainment of advanced courses (Glasman and Biniaminov, 1981). School quality is not predicted by the numbers of teachers with advanced preparation. Sustained effort to use staff development in the context of general school reform has been lacking. As Guskey (1986, p.5) put it, "Nearly every major work on the topic of staff development has emphasized the failings of these efforts." Stallings and Krasavage (1986) argue that even highly directed training in specific instructional skills conducted over a period of several years did not result in sustained changes in teaching behaviors. Slavin (1989) placed the ineffectiveness of staff development in the larger context of fads in education. Fenstermacher and Berliner (1985) have written about the lack of evaluation models for understanding the effectiveness of staff development, and have proposed a way to begin to do so. Finally, data from the *METROPOLITAN LIFE SURVEY OF THE AMERICAN TEACHER* (1986) indicate that about 75 percent of teachers wished to influence the design and conduct of staff development programs, but only about 30 percent felt that they did so. A recent summary of staff development says that it is an "enterprise that is fragmented, not frequently engaged in on a continuing basis by practitioners, not regarded very highly as it is practiced, and rarely assessed in terms of teacher behavior and student learning outcomes." (Howey and Vaughn cited in Guskey, 1986, p.5)

It is quite difficult to imagine what kind of evidence would address the general question of the level of success of staff development efforts in the United States. At the most abstract level perhaps staff development has been successful. Over the years teachers have been able to adapt technique and curricula to changes in policy mandates. If schools have changed at all in fifty years, one must admit the possibility that staff development has contributed to these changes. In addition, policy makers must see some benefit in staff development because it continues to receive funding, and policy makers continue to worry about its content, quality, and form.

At more concrete levels, the evidence is much less certain. Because staff development is so pervasive, no large-scale studies of it effects have been done. The assessments of Teacher Corps and the Teacher Centers did not prove compelling enough to sustain them. Cuban (1984) argued that "... over nearly a century, the data show striking convergence in outlining a stable core of teacher-centered instructional activities in the elementary school and, in high school classrooms, a remarkably pure and durable version of the same set of activities." (p.238)

Aside from the effect of staff development efforts, their quality is a major issue. Shoddy work is tolerated perhaps because teachers have come to expect little from staff development, the "proofs of purchase" continue to be available and no professional standards are available to assess the activity. The system is so diffuse that word-of-mouth assessments may or may not affect subsequent provider behavior. Often enough, evaluation is conducted against teacher perceptions of usefulness or likability, but almost never against a standard having to do with school improvement. Finally, it is probably fair to say that entertaining presentations on "hot" topics get far better marks from teachers than the content or consequence would justify. The profession seems to have agreed tacitly that since staff development is not to be taken seriously anyway, great variations in quality are tolerable.

A second serious problem has to do with quantity. No evidence exists to allow a sensible policy decision about the amount of staff development needed to accomplish any given purpose. This is so in part because activity and purpose are so seldom connected. Private providers charge hourly rates, so the amount of staff development is a function of a district's willingness to spend. The provider simply matches the quantity of the service to the contract price. Universities and colleges operate on the basis of credit hours, with course material tailored to fill up the number of contact hours required to satisfy a definition of a credit hour. No standards exist to help define how much a person might expect to learn in a one or three credit hour course. The matter rests almost entirely with the faculty member teaching the course.

A third problem is one of distribution. Teachers in urban areas have choices and exposures that teachers in remote areas lack. Because staff development delivery is labor intensive, teachers in remote areas must often travel great distances, rely on local talent, or engage with a variety of "non-traditional" delivery systems.

Despite research of varying focus and quality, including perhaps the largest single effort to assess results (Little et al., 1988), staff development efforts continue and expand based on the assumption of benefit to the public. The system rumbles on, unchecked and effectively unexamined.

## POLICY ALTERNATIVES

If the central argument of this article is sound--viz., that current models of staff development neither deliver, nor promise to deliver, predictable increases in school quality-- some obvious policy alternatives can be examined. The first, of course, is to do nothing. Powerful vested interests would support the option. Row and column salary schedules are quite attractive to teacher associations. Movement across and down is relatively painless, with staff development providing easy mechanisms for enhancing life-time earnings. Providers of all sorts benefit in many ways. New ideas, for good or naught, do get disseminated. Participants may benefit in other than economic ways. Thus, in the absence of documented harm, and with undocumented expenditures, the political cost of making major changes may be too high.

A second option would be to abandon the basic assumption that staff development makes any difference to anyone, and get out of the business with public funds. The elimination of public funding for staff development might free up substantial dollars for other efforts to increase student performance. But this option is as unattractive as the first is attractive, and for the same reasons. Substantial numbers of individuals and groups benefit from the current system. Consequently, it is an unlikely choice.

A third option is for state and local education agencies to develop policies which increase the possibilities for successful staff development. To do so, however, requires that policy decisions be informed by an understanding of the alternative forms that such programs can take and how these are related to adopted goals. Much in a prescriptive nature has been written about successful delivery systems (Dilworth and Imig,1995; Howey, Bents and Corrigan, 1980; McKenzie 1980; Academy for Educational Development, 1985; Hall, 1986; Fielding and Schalock, 1985).

However, two prior considerations modify the structure of a staff development system; the content of the training to be provided and the methods of program delivery.

### *Content*

In content, staff development programs provide some combination of technical and interpersonal or organizational skills. Technical skills include subject matter expertness and pedagogical techniques along with such ancillary bodies of knowledge as child development, student assessment, and classroom management. These skills, critical in the development and effective implementation of instruction, inform the selection of materials, modification of instruction to meet the needs of diverse student groups, identification of alternative learning experiences, adaptation of lesson plans to changing classroom contexts, and so on. They significantly affect day-to-day classroom operations (Mitchell, Ortiz and Mitchell, 1983). Although colleges and universities are expected to develop these skills in preparation programs, beginning teachers cannot be expected to have mastered them, as Berliner (1986) has shown. Staff development, thus, may have a legitimate role to play in continued skill development.

Beyond the inculcation of improved craft skills, staff development programs might help teachers understand and use organizational skills needed to work effectively with other adults. These skills include learning to participate in decision-making groups, to assess and plan for overall school improvement, and to interact with groups of parents. Put more generally, these are the skills required to work as colleagues with other adults in a professional setting (Blankenship, 1977). Modern schools are not simple organizations as portrayed in the folklore of American education.

While much of the work of schools continues to be done in a setting where one teacher works with small groups of students on simple and standardized lessons, this image obscures as much as it reveals. Schools are complex social enterprises filled with different roles for students and teachers. Strong and frequently divergent points of view shape the behavior of professional educators at all levels. Moreover, differences in community values and social systems subject the schools to competing interests.

The complex social order of the modern school creates even more need to master interpersonal or organizational skills. Mentoring or coaching, for example, requires a set of skills that teachers rarely develop during their initial preparation. Mentors must be able to assess the behavior of a colleague, counsel with the person and provide suggestions for improvement while, at the same time, retaining a peer relationship. The introduction of school- site councils requires that teachers learn an array of group process skills. Once again, these are not skills which routinely appear in preparation programs.

### *Delivery*

In addition to emphasizing different types of training content, staff development programs differ significantly in form. From a policy perspective, the most important factor in the form of delivery is whether training opportunities are directed to individual teachers or to groups of teachers with common work responsibilities. The first approach is more common, used when the goal is to improve individual performance by allowing teachers to identify their own needs and preferences and to select training opportunities without reference to others in the same school organization. This is the modal form of delivery in the current model.

The alternative, the work-group approach, has grown more popular in recent years. Here the main goal is to strengthen institutional capacity by encouraging teachers to think of staff development as an integral part of an overall school or district improvement program. It is usually delivered in the form of workshops or seminars focused on school site, grade-level, or subject matter problems that require coordinated responses from both teachers and administrators.

### *Content/Delivery Models*

In combination, the content and delivery variables define four models of staff development for teachers. As shown in Table 1, these four models define the parameters of a system.

**Table 1**
**Models of Staff Development**

| | | FOCUS OF DELIVERY | |
|---|---|---|---|
| | | Individual | Work Group |
| CONTENT | Pedagogy/Instruction | Instructional Enhancement | Program Development |
| | Organization Leadership | Professional Leadership | School Improvement |

Instructional enhancement is the traditional mode, and is served by staff development programs that combine technical skill development with a focus on delivering services to individual teachers. Skills such as new instruction methods, classroom management, diagnosis of student learning problems, motivation techniques and the use of curricular materials are typically taught in this way.

The lower left cell describes staff development designed to enhance professional leadership. The content of training shifts from technical to organizational skills, although the focus continues to be on individuals. Department heads, mentor teachers, team leaders and master teachers are obvious participants. Each needs to know how to function effectively with other adults and to operate within complex social roles that are not ordinarily contemplated, much less developed, during preparation or in early career.

Improved functioning for groups of teachers working on program development tasks is the focus of activity in the upper right cell. It is one thing for a single teacher to plan lessons for a year, but quite another to establish the scope and sequence of science or mathematics instruction for a particular grade level in an entire school. Teachers must learn high levels of technical skills, not generally applicable in individual classrooms. Textbook assessment, curriculum alignment, program evaluation, and student assessment models are examples of these sets of skills. Their application is conditioned by the context of school and district-level decisions regarding emphases and directions. As decisions of the group affect its various members, teachers participate as members of work groups rather than as individuals.

The activity represented in the lower right cell has as focus overall school improvement. In order to make schools more robust learning places teachers combine their personal skills with organizational processes that can only be acquired and exercised in a work-group setting. Even the best teachers will be less than optimally effective if they succumb to intra-faculty squabbles over teaching methods, coordination and cooperation, or school directions. This cell represents much of what is required to bring about collaborative, school-based, change. Taken together, the processes of genuine change in a school are quite involved (Dillon-Peterson 1981) and require sophisticated interaction skills.

The four types of staff development are available to policy makers. But because these models are designed to accomplish different ends, the links to policy objectives need to be made clear.

## EVIDENT POLICY CHOICES

Policy choices are statements of value. As such they rest on both desired ends and on assumptions about the relationships between ends and means (Marshall et al., 1985). In the absence of reliable data about what really works to make schools better, policy makers operate from what they believe will work or from ideas which they believe will satisfy their own self interests. The choice of staff development models, in turn, rests on those beliefs, and each staff development model has behind it a different assumption about how to reform schools. It makes sense for policy makers to be clear about their assumptions concerning school reform, because the choice of staff development emphases can be made consistent with them.

If policy makers believe that the primary tool for improving education is to hold teachers strictly accountable for performance, then the Instructional Enhancement model of staff development is appropriate. This is so because teacher accountability policies make the logical, but narrow, assumption that the best way to make better schools is to make better teachers. This is to be accomplished by tougher performance evaluation standards and, sometimes, by linking performance to compensation. Doing so might encourage teachers to look beyond the "proof of purchase" utility of staff development programs and to concentrate on experiences that they thought would increase the probability that training would improve chances for positive evaluation or increased compensation.

If policy makers believe that the key to school improvement lies with creating new teacher work roles, the Professional Leadership model of staff development is preferred. This policy strategy assumes that overall school improvement will result if teachers accept more differentiated job responsibilities and make unique contributions to the instructional program. In effect, this strategy attempts to increase the general density of instructional leadership in a school.

If policy makers believe that professionalizing teaching is a precursor to school improvement then the staff development strategy of choice should be that of Program Development. The key assumption of this policy strategy is that a professionalized workforce in the schools will find more effective ways of cooperation and collaboration in school program design and implementation. Professionalization, the argument goes, means that teachers will become intimately involved in the design and assessment of programs. As a result, they will accept more responsibility for the quality of their implementation and will work closely with their colleagues to insure that all students are given appropriate opportunities to learn. Staff development is concentrated on technical aspects of the instructional process not ordinarily exposed to teacher control.

Finally, if policy makers believe that improving education is more likely if schools are restructured, the staff development strategy of School Improvement may be attractive. The intent of the strategy is to transform schools into cooperative learning communities in which student needs become paramount by altering decision-making procedures, organization structures, and the distribution of authority and responsibility. The needs of the school, as identified by teachers, students, parents and local administrators determine the scope and nature of the staff development work undertaken. In this environment, staff development is a continuous and central element of life-not a special set of programs or activities.

As argued, the choice of staff development model can and ought to be linked to beliefs about central strategies for school reform. Depending on assumptions policy makers make about what will work, staff development models will vary. It is not clear which, if any, of these is most successful, but it is clear that each is designed to accomplish different ends. If policy makers mix ends and means, as they do now, the results are unlikely to be different from the current muddle in which staff development is provided.

## IMPLICATIONS FOR THE STAFF DEVELOPMENT MARKETPLACE

So far I have argued that current staff development policy and implementation is flawed on two counts; little deliberate connection is made between the presumed purposes of staff development and the various means by which it is carried out, and the marketplace for staff development is an imperfect consumer market in which "proofs of purchase" can substitute for utility. I have argued as well that it is possible to articulate four distinct models of staff development, each anchored in a distinct set of assumptions about how to improve schools. In the next section I explore how each of the four school improvement strategies, and its associated staff development model, has a potential effect on the marketplace for staff development. No claim can be made that these consequences are likely, since a prior claim was made that "doing nothing" is the likely policy choice. But they are interesting speculations.

### *Instructional Enhancement Models*

The accountability strategies, so prevalent in recent reform efforts, may already have begun to shift the market system away from open choice and high levels of teacher discretion toward authoritative definitions of required technical skills. Generally accepted standards of good practice are being incorporated into standards for staff training and evaluation in many states, as are the knowledge and experience seen as the basis for their mastery. The most likely consequence is that the variety of staff development programs and activities will be reduced, sharpening the focus and intellectual definitions of teaching. Publication of TOWARD HIGH AND RIGOROUS STANDARDS FOR THE TEACHING PROFESSION (National Board for Professional Teaching Standards, 1989), and subsequent assessment of individuals by the Board is an obvious first step toward establishing uniform standards of teacher performance.

In the long run, one can hope that teaching will become a more rigorous field of study. Under accountability pressure, teachers might be expected to seek high quality staff development programs more explicitly linked to the skills required for positive appraisal, salary advances and job retention. As the Carnegie Task Force (1986) put it, there is "no reason to perpetuate a system of continuing education that determines teacher compensation on the basis of credits earned after becoming a teacher. Compensation should be based on proven competence, not time in the chair" (p.77).

Carefully developed accountability policies might be expected to have a second important effect on the staff development delivery system. By establishing performance-based criteria for certification and recertification, accountability policies encourage the development of richer and more comprehensive teacher assessment practices. In addition to standardized tests of pedagogical and subject-matter knowledge, recent accountability proposals include requirements that teachers prepare a work portfolio containing such artifacts of competence as lesson plans, teacher-made tests, instructional materials, or videotapes of teaching. Staff development which cannot, in some demonstrable way, contribute to a richer portfolio might become unattractive to potential consumers.

One possible result will be new staff development vendors. Private coaching schools aimed at facilitating the acquisition of needed knowledge and skills would have a natural market if the financial rewards approached the rewards available in other fields. Publicly supported colleges and universities might not compete vigorously in this market. Prestige law schools and private cram schools exist side by side. If repeated in education, university based schools of education might concentrate on pedagogical theory and research, leaving specialized skill development to other vendors.

### *Professional Leadership Models*

These strategies could lead to greater specialization in the delivery of staff development services. To the extent that differentiated staffing in schools becomes a reality, programs will become available to support mobility into various specialized roles within the school. Training for mentor teachers, peer coaches, curriculum developers, department chairs and other new roles could follow the well-developed pattern of specialized training for school counselors, reading specialists, and school administrators.

Training in the new roles of teacher leadership is likely to be more on-the-job than otherwise, because such jobs are likely to be filled by persons who are chosen by their peers. Colleges, universities and private vendors will undoubtedly develop packages of short-term training which incorporate specific skill sets. The market is likely to be segmented and the purchasers are more likely to be districts than individuals.

### *Program Development Models*

State policy strategies supporting professionalization can be expected to have a different effect on the staff development market system. The responsibility for program development, implicit in these models, will rest in large part with groups of teachers working together. It is at least possible that individual teachers will become expert in certain areas and will be able to coach their colleagues. Within schools and school districts we may see an increasing "in-house" capacity for staff development and the adoption of locally designed "trainer of trainers" models. Outsiders might be brought in for purposes of helping design staff development systems, but the direct services may well be provided by local talent. In addition, state department of education employees can serve as useful technical assistance providers. The net effect of this may be to reduce course taking by teachers and to increase on the job training provided by school district employees.

### *School Improvement Models*

The fourth general model can be expected to have the most profound effect on staff development marketing. In order to support restructuring, staff development activities have to be placed in the hands of local school or district leaders, and staff training has to be merged with school program and policy development so that skill enhancement is parallel to shifting responsibilities of staff.

Within this framework, control over staff development resources needs to be linked to overall school leadership responsibility. Whether placed in the hands of teachers or administrators, leadership for school restructuring will need to combine new organizational designs with new systems of resource and authority allocation. In order for redesigned schools to work, staff development resources will have to be focused directly on helping all members of the organization make the transition and become contributors within the new structural framework. Staff development, therefore, will have to give up its emphasis on service to individuals and become an integral part of organizational planning and development.

## FROM A CONSUMPTION TO AN INVESTMENT MODEL

The current state of staff development is in disarray and driven by undesirable market conditions. By connecting staff development to school improvement, the staff development market can be changed from a largely unregulated consumer market to one in which quality is demanded by persons who view staff development as an investment decision rather than a consumption decision. In this model, return on investment becomes the decision criterion, and the rate of return will be indicated by the level of progress of school improvement. Such a model will force higher quality experiences.

If return on investment were to become the primary decision criterion, two consequences become apparent. The first consequence would be substantial reduction in the cafeteria-like offerings now in the market and possibly an end to the proliferation of "courses" offered by colleges and universities, county offices of education, state departments of education and school districts. Without the sure return on investment provided by the proof of purchase, teachers might simply stop accumulating credits. (This assumes, of course, that policy makers have sufficient motive to abandon current row and column salary schedules.) Aside from some dislocations in the workforce of providers, the result might be a substantial reduction in public subsidy.

In addition, relieved of pressures to offer courses and workshops for the convenience of the "credit collectors," universities and colleges might give serious attention to constructing degree programs of rigor and intellectual integrity. Teachers might then choose to take advanced degrees because of clear evidence that doing so would improve their work performance or their intellectual quality of life. They might also demand much higher levels of performance by faculty since the teachers would have to risk the cost of tuition and fees against no clear return on investment.

The second consequence may be a shift in the structure of providers, with the dominance of colleges and universities giving way to entrepreneurs. New providers and new technologies for delivery could develop. College and university faculty in education might begin to differentiate the unique roles of university study, and a degree in education may come to have some common meaning as staff development activity is assumed by other agencies. This might mean a reduction in the size of education faculties, though normal attrition would offset any sudden dislocations.

An investment market in which anticipated return would drive up the quality of offerings and would be linked to strategic policy choices about school reform leaves unanswered the questions of "Who invests?" and "Who benefits?" At least two general answers are available. The first is that the primary beneficiary is the teacher, and thus the teacher should make the major investment. The argument makes most sense if accountability models or leadership development models are the strategies of choice for school improvement. Because the teacher will benefit directly by job retention, higher pay, or increased job responsibility, the teacher should pay for the skills, as private practice professionals do.

But this argument has two flaws. The first is that the level of return on investment is quite low for teachers. A base salary in the $20,000-$30,000 range is not comparable to the salaries of private practice professionals. In addition, publicly financed institutions typically cap the salaries of even the highest performing individuals. Thus, the decision by a teacher to invest is bounded by narrowly defined returns.

The second flaw is that teachers work in public bureaucracies and do not have full discretion to practice their craft. They are expected to accept institutional goals and constraints. Consequently the returns to them are modified by institutional demands and interventions. Newly acquired skills may not be used if they conflict with institutional policies, procedures, and cultures.

The difficulties with placing the decision to invest with the teacher suggest that school district officials should make major investment decisions. Currently, teachers make the major consumption decisions and the cost of those decisions are passed through to the public, with no apparent relationship to improvement. Were school districts to take seriously the investment model, decisions about participants, content, cost, delivery and the rest would be made only after consideration of the underlying question of expected return. Then justification for public subsidy could be stated and debated. At present the debate over return is not held because market mechanisms deflect such questions.

School districts can decide the mix of services, identify providers, assess results, and determine, finally, the quality of available services. State and federal roles would include monitoring school district decisions and suggesting or mandating alternative strategies or providers. If teachers chose to study for degrees or to buy experiences outside those sponsored and paid for by districts, they would do so as private investors with no guarantee of return on their investments. Thus, the proof of purchase would disappear as a measure of utility.

## CONCLUSION

Staff development has had a spotty record in American education. Most thoughtful persons will at once agree that it is a necessary activity and that it is unsatisfactory in its current form. By linking staff development to strategies for school improvement, policy makers can rethink the purposes, structures, and content of future efforts.

The current lack of focus in staff development policy derives from the disjunction of activity and purpose and the domination of an imperfect and inappropriate consumer market. It has been shown that the goals, delivery, and content of staff development can be linked differentially to strategies for school improvement. By doing so, policy-makers can change the market to one driven by investment decisions, can raise the overall quality of experiences, and build a base for clearly assessing the returns on investment.

## References

Academy for Educational Development. (1985) *TEACHER DEVELOPMENT IN SCHOOLS: A REPORT TO THE FORD FOUNDATION*. New York: Academy for Educational Development.

Berliner, D. (1986) In pursuit of the expert pedagogue. *EDUCATIONAL RESEARCHER*. August/September. 5-13.

Blankenship, R.L., (ed.) (1977) *COLLEAGUES IN ORGANIZATION: THE SOCIAL CONSTRUCTION OF PROFESSIONAL WORK*. New York: John Wiley and Sons.

Carnegie Task Force on Teaching as a Profession. (1986) *A NATION PREPARED: TEACHERS FOR THE 21ST CENTURY*. New York: Carnegie Forum on Education and the Economy.

Cuban, L. (1984) *HOW TEACHERS TAUGHT*. New York: Longman.

Dillon-Peterson, B.(ed.) (1981) *STAFF DEVELOPMENT/ORGANIZATION DEVELOPMENT*. Alexandria, VA: Association for Supervision and Curriculum Development.

Dilworth, M. and Imig, D. (1995) Professional Teacher Development. *THE ERIC REVIEW.* 3(3). Winter 1995. 5-11.

Fenstermacher, G. and Berliner, D. (1985) Determining the value of staff development. *THE ELEMENTARY SCHOOL JOURNAL.* 85(3). 282- 314.

Fielding, G.D. and Schalock, H.D. (1985) *PROMOTING THE PROFESSIONAL DEVELOPMENT OF TEACHERS AND ADMINISTRATORS.* Eugene, OR: Center for Educational Policy and Management, University of Oregon.

Glasman, N.S. and Biniaminov, K. (1981) Input-output analyses of schools. *REVIEW OF EDUCATIONAL RESEARCH.* 51. 509-41.

Guskey, T.R. (1986) Staff development and the process of change. *EDUCATIONAL RESEARCHER.* May. 5-12.

Hall, G.E. (1986) *BEYOND THE LOOKING GLASS: RECOMMENDATIONS AND CRITICAL WARNINGS FOR TEACHER EDUCATION PRACTITIONERS, POLICY MAKERS AND RESEARCHERS.* Austin, TX: The Research and Development Center for Teacher Education, The University of Texas at Austin.

Holmes Group. (1986) *TOMORROW'S TEACHERS: A REPORT OF THE HOLMES GROUP.* East Lansing, Michigan.

Howey, K.R., Bents, R. and Corrigan, D., (eds.) (1981) *SCHOOL- FOCUSED INSERVICE: DESCRIPTIONS AND DISCUSSIONS.* Reston, VA: Association of Teacher Educators.

Lanier, J.E. and Little, J.W. (1986) Research on teacher education. Wittrock, M. (ed.) *HANDBOOK OF RESEARCH ON TEACHING* (Third Edition). New York: Macmillan. pp. 527-69.

Little, J.W., Gerritz, W., Stern, D., Guthrie, J., Kirst, M. and Marsh, D. (1988) *STAFF DEVELOPMENT IN CALIFORNIA.* San Francisco: Far West Laboratory for Educational Research and Development.

McKenzie, F. (1980) Staff development as a means of improving schools. National Council of States on Inservice Education. *PROVIDING LEADERSHIP FOR STAFF DEVELOPMENT AND INSERVICE EDUCATION.* Syracuse, N.Y.: National Dissemination Center, Syracuse University School of Education.

Marshall, C., Mitchell, D., Wirt, F. (1985) Assumptive worlds of education policy makers. *PEABODY JOURNAL OF EDUCATION.* 62(4). 90-116.

Metropolitan Life Insurance Company. (1986) *THE AMERICAN TEACHER,* 1986. New York: The Metropolitan Life Insurance Company.

Miller, B., Lord, B., and Dorney, J. 1994. *SUMMARY REPORT. STAFF DEVELOPMENT FOR TEACHERS. A STUDY OF CONFIGURATIONS AND COSTS IN FOUR DISTRICTS.* Newton, MA: Education Development Center.

Mitchell, D. (1986) State policy strategies for improving teacher quality. *POLICY BRIEFS.* San Francisco: Far West Laboratory for Educational Research and Development.

Mitchell, D.E., Ortiz, F.I. and Mitchell, T.K. (1983) *WORK ORIENTATION AND JOB PERFORMANCE: THE CULTURAL BASIS OF TEACHING REWARDS AND INCENTIVES.* A report to the National Institute of Education.

National Board for Professional Teaching Standards. (1989) *TOWARD HIGH AND RIGOROUS STANDARDS FOR THE TEACHING PROFESSION.* Detroit: National Board for Professional Teaching Standards.

Slavin, R. (1989) PET and the pendulum: Faddism in education and how to stop it. *PHI DELTA KAPPAN.* 70, June. 752-758.

Stallings, J. and Krasavage, E.M. (1986) Program implementation and student achievement in a four-year Madeline Hunter Follow- through Project. *THE ELEMENTARY SCHOOL JOURNAL.* 87(2). 117-138.

Stout, R. and Wigand, R. (1982) Education. Karnig, A. and Hall, J., (eds.) *THE IMPACT OF THE NEW FEDERALISM ON ARIZONA.* Tempe, AZ: The Morrison Institute for Public Policy, Arizona State University.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Mark E. Fetler
*fetlerctc.aol.com*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 3** **February 26, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

# Making Molehills Out of Molehills: Reply to Lawrence Stedman's Review of *The Manufactured Crisis*

**David C. Berliner**
**Arizona State University**

*berliner@asu.edu*

**Bruce J. Biddle**
**University of Missouri**

*psybiddl@mizzou1.missouri.edu*

**Abstract:** Berliner and Biddle answer Lawrence Stedman's review of their book The Manufactured Crisis, which was published in the Education Policy Analysis Archives as Volume 4, Number 1, 1996.

Throughout his term as founding editor of "Contemporary Psychology," Edwin G. Boring insisted that the basic tasks of the responsible reviewer are to portray with honesty the intentions of authors and to assess carefully whether those intentions are realized in their writings.

Unfortunately, Lawrence Stedman (1996) does not honor such laudable tenets in his so-called "review" of our book, THE MANUFACTURED CRISIS, appearing in Education Policy Analysis Archives, 4(1). Instead, Stedman chooses to ignore both the intentions that we stated clearly in our book and the vast bulk of what we actually wrote about in its eight chapters. Worse, he asserts falsely that our book was based on four "sweeping claims" and then attacks us because the analyses with which we supposedly supported these claims were "deeply flawed and misleading."

In fact, these so-called "sweeping claims" referred to materials covered in but a portion of our second chapter. Further, two of Stedman's concerns about our "sweeping claims" misrepresented what we had written, and the other two state positions with which Stedman agrees and are abundantly supported by the evidence he himself cites. In short, Stedman has written a review that is uninformative, disingenuous, and as will soon become clear, trivial.

Stedman has not succeeded in even making a mountain out of a molehill--all that was accomplished was to make molehills out of molehills.

## WHAT WE WROTE ABOUT

Since Stedman does not bother to tell readers what we actually wrote about in THE MANUFACTURED CRISIS, we should begin by doing so. We began our book by noting that throughout most of the Reagan and Bush years, the White House led an unprecedented and energetic attack on America's public schools, making extravagant and false claims about the supposed failures of those schools, and arguing that those claims were backed by "evidence." To illustrate, in 1983 the White House released a widely-touted brochure, "A Nation at Risk," claiming (among other things) that the "average achievement of high school students on most standardized tests is now lower than 26 years ago when Sputnik was launched." This claim made an assertion about factual matters, but somehow no evidence was cited in "A Nation at Risk" to support it, nor could any have been given since it was false.

Again, in 1989 John Sununu was to claim that Americans "spend twice as much [on education] as the Japanese and almost 40 percent more than all the other major industrialized countries of the world," and George Bush (the "Education President") was to intone that our nation "lavishes unsurpassed resources on [our children's] schooling." These claims were equally untrue. Other damaging claims made by the White House during these years argued: that American schools "always" look bad in international comparisons of achievements; that educational expenditures are not related to school achievements and that additional investments in education are "wasted"; that because of inadequacies in our schools, American industrial workers are non- productive; and that the typical private school out-achieves the typical public school when dealing with similar students. These and other false claims, designed to weaken Americans' confidence in their public schools, were all said to be backed by "evidence," although somehow the "evidence" in question was often only hinted at.

This attack was led by specific persons--whom we named in our book--and created myths about education that were sometimes backed by no evidence at all, sometimes supported by misleading analyses of inappropriate data, and sometimes aided by the deliberate suppression of contradicting information. No such White House attack on public education had ever before appeared in American history--indeed, even in the depths of the Nixon years the White House had not told such lies about our schools. Since the attack was well organized and was led by such powerful persons--and since its charges were shortly to be echoed in other broadsides by leading industrialists and media pundits--its false claims have been accepted by many, many Americans. And these falsehoods have since generated a host of poor policy decisions that have damaged the lives of hard-working educators and innocent students.

In our book we labeled this attack "The Manufactured Crisis" and detailed:

- the abundant evidence that contradicts its major myths;
- the likely reasons for its appearance in the Reagan and Bush years;
- the ways in which the "reform" proposals associated with this attack would be likely to damage America's public schools;
- the real and escalating social problems faced by our country and its schools, that leaders of the attack had but little interest in solving; and
- what can be done today to help solve the real problems of our schools.

As this brief summary suggests, our book was designed to cover a good deal of material. In it we also tried to write not a scholarly treatise but rather a work that could be read by the wide audience of educators, policy-makers, parents, and citizens in our country who are truly concerned about education today. However, these intentions are neither noted nor assessed by Stedman, so readers will have to read THE MANUFACTURED CRISIS themselves to find out whether or not we succeeded in accomplishing them.

## DISINGENUOUS CHARGES

So much for Stedman's sins of omission. What about those he committed? In his lead paragraph, Stedman asserts that our book made four "sweeping claims" about American educational achievement and implies that these constitute the core of our arguments in TMC. This is nonsense, of course. The four "claims" in question do not portray the major themes of our book. Rather, they focus only on narrow issues of student achievement that are dealt with in but part of our second chapter.

In addition, two of the supposed "sweeping claims" challenged by Stedman misrepresent what we actually wrote. One asserts that we had concluded, "today's students are 'out-achieving their parents substantially' (p. 33). " This quote was taken out of context. In one short sub-section of Chapter Two we reviewed longitudinal evidence from commercial tests of achievement such as the Iowa Test of Basic Skills, the California Achievement Test, and the like. Citing evidence originally developed by Linn, Graue, and Sanders (1990), we noted that for some years average scores earned on these tests have been creeping upwards and that the test developers have regularly had to recalibrate these tests in order to make certain that the typical student again scores at the fiftieth percentile rank for the subjects assessed. Commenting on this brief review, we wrote "So, if commercial tests were not recalibrated, virtually all of them would show that today's students are out-achieving their parents substantially" (p. 33), and this sentence was the source of Stedman's misleading quote.

We never claimed that equivalent effects have appeared in the more extensive evidence from non-commercial tests of student achievement, nor did we state any general conclusions about today's students out- scoring their parents in school achievement anywhere in our book. So Stedman's assertion that we had made such a "sweeping claim" is not so. In fact, we were actually quite cautious in what we claimed about the achievements of students and their parents.

But while we are on the subject, related thoughts may be worth mentioning. As we noted in TMC, IQ test data from over a dozen industrialized nations show that today's children are about one standard deviation ABOVE their parents in measured intelligence, with the growth primarily in the decontexualized, abstract, problem-solving parts of the tests (sources cited in our book). Additionally, when one looks at more than 20 "then" and "now studies of student achievement--reviewed previously by Stedman himself in his studies of literacy in the U. S.!--almost all the results show that the students taking the test "now" outscore the students that took the test "then." So while we were actually cautious in our book, and did not make the "sweeping claim" assigned to us by Stedman, the data suggest that such a claim might actually be made!

In addition, Stedman asserts that we made another "sweeping claim," that "the general education crisis is [merely] a right-wing fabrication," although he provides no citation to justify this charge. Again, this misrepresents what we wrote. Rather, we devoted an entire chapter in our book to a careful analysis of the social origins of The Manufactured Crisis, and in it we pointed out that this episode in American history reflected MANY causes. It is certainly true that right- wing ideologues gained access to the White House with the election of Ronald Reagan, and in our book we detailed their influence on White House education policy. But school-bashing has been a popular indoor sport in America for years, and White House critics of the schools would not have gotten away with the lies and distortions of evidence they promoted had Americans not also been worried about unresolved problems in our society and its public schools, and had their efforts not been supported by industrial pronouncements and media irresponsibility. Thus, by reducing our careful analysis to a political slogan, Stedman has seriously distorted what we wrote in TMC.

So on two of our "sweeping claims," Stedman misrepresented us. As we shall see below, however, Stedman states that he generally agrees with the other two "sweeping claims" he correctly assigns to us. The additional evidence he cites provides no reason to question our interpretations of the data. We turn now to these issues.

## CREATING MOLEHILLS, PART ONE--THE MYTH OF DECLINING TEST SCORES

The first of the "sweeping claims" which Stedman accurately assigns to us concerns the myth of declining test scores. After reviewing evidence from many sources, we DID write, "standardized tests provide no evidence whatever that supports the myth of a recent decline in the school achievement of the average American student" (p. 34). Moreover, Stedman states that he agrees with this claim, writing, "Berliner and Biddle are generally right that achievement has been stable," and again, "the best that can be concluded is that this generation of students generally performs about the same as earlier ones." So--to paraphrase a recent hamburger commercial--where's the beef?

Stedman goes on to complain that we had not reviewed even more evidence on the issue, cites various materials that HE had reviewed in previous publications, and implies that somehow these additional materials would cause one to rethink or possibly to revise the claim we had made (and with which he clearly agrees). But would additional insights have been gained had we added these extra materials to a chapter that was already overly long? To answer this question, let us scan the evidence alluded to by Stedman.

For openers, Stedman complains about our portrayal of NAEP results. He writes that "high school students' NAEP civics scores, for example, dropped substantially between 1969 and 1976 and have been slipping ever since." But is this true, and is it a substantive matter? Evidently not. NCES's "The Condition of Education, 1991" noted that no statistically significant differences appeared in average NAEP civics scores between 1976, 1982, and 1988 for either 13-year-olds or 17-year-olds (1991, pp. 143, 144). One data set showed slight gains, the other showed slight losses, but evidently neither of these "trends" mattered.

Stedman also claims that "[NAEP] science scores also fell during the 1970s and have only partly rebounded," but again is this true, and is the matter substantive? Let readers judge for themselves. Average NEAP science scores for the years 1970, 1973, 1977, 1982, 1986, 1990 and 1992 were: For 9-year-olds 225, 220, 220, 221, 224, and 229 and 231; for 13-year-olds 255, 250, 247, 250, 251, 255 and 258; and for 17-year- olds 305, 296, 290, 283, 288, 290, and 294, respectively (National Center for Educational Statistics, 1994, p. 56). In short, Stedman's judgment about science scores is simply wrong! Over 22 years, two of the three age groups studied actually showed slight GAINS during this period, but the most reasonable interpretation of the science data is again one of general stability over time.

Stedman also writes, "in the early 1990s, younger students' NAEP reading and writing performance slipped." Again, let readers judge the issue. Reading scores reported for 9-year-olds over seven administrations of the NAEP covering 21 years were: 208, 210, 215, 211, 212, 209, and 210, respectively (National Center for Educational Statistics, 1994, p. 50.). Thus Stedman's interpretation of the data is once again wrong! He sees a decline in reading scores when he should be seeing remarkable consistency of scores over time. In addition, the NAEP writing test seems to have been administered four times between 1984 and 1992, and the following average scores were earned: for Grade 4-- 204, 206, 202, and 207; and for Grade 8--267, 264, 257, and 274; (National Center for Educational Statistics, 1994, p. 52). As before, Stedman's interpretation seems to be in error. It is difficult to understand how Stedman could misread such stable data sets and conclude that they indicate "slippage." (Curious readers may check the NAEP data for themselves. They appear in all recent editions of the CONDITION OF EDUCATION.)

For some reason, Stedman also chooses to complain about our review of SAT evidence. He challenges our conclusion that the notorious, so- called "decline" in SAT scores in the late '60s and early '70s was largely generated by sharp increases in the range of students opting to take the test, asserting that we had ignored his published demonstration that demographic changes in test takers explain "much, but not all" of this decline in SAT scores. Two crucial points are relevant to this complaint. First, how could Stedman or anyone else possibly know whether demographic changes do not explain all of the notorious SAT "decline" since MANY important demographic characteristics of students are never measured and thus cannot be entered into analyses concerned with the shifts in SAT scores? But more importantly, in the process of issuing his complaint, Stedman utterly ignores the point often made by other scholars, and repeated forcefully in TMC, that aggregate SAT scores are NOT valid for judging the achievements of school districts, states, or the nation as a whole because they are not based on random samples. So this complaint turns out to be a true tempest in a teapot. (Despite which, some readers may continue to wonder about other possible reasons for the SAT "decline." A plausible hypothesis is offered in Note 1.)

In addition, Stedman challenges another of our conclusions that he does not bother to document. Based on disaggregated evidence from both SAT and NAEP scores, we asserted that the overall achievements of minority students have recently been slowly improving in America. In apparent contradiction, Stedman states that we had ignored SAT evidence showing "minority verbal declines in the late 1970s and late 1980s." But it is far from clear that these putative "declines" were substantive; the evidence for these putative "declines" in SAT scores was matched by more representative national data from the NAEP that showed large gains in minority reading scores between 1971 and 1992 ( National Center for Educational Statistics, 1994, p. 50 ); and once more the point made by Stedman does not contradict the general conclusion we wrote about in TMC. Thus again, there is less here then meets the eye.

Finally, Stedman accuses us of writing a "selective" review of the work of Linn, Graue, and Sanders (1990) on commercial tests: failing to report data from the SRA; failing to report data that Linn et al. had generated on high school achievement; and failing also to note their "worries" that recent gains in commercial test scores might have reflected school districts' repeated use of the same tests rather than genuine student improvement. Let us put these concerns to rest.

- Regarding the SRA issue, the data reported by Linn et al. are complex and mixed, and we judged that they required too much explanation to warrant their inclusion in a book designed for general readers--but those data do NOT contradict the interpretation we gave (see Note 2).
- Regarding the high school issue, we chose again to leave the data out because academic achievement growth in basic subjects seems to be limited at the high school level (see Coleman, Hoffer, & Kilgore, 1982, for example) and because Linn et al. did not report high school data for the CTBS and the ITBS--but again, the high school evidence does NOT contradict the conclusion we stated. (in fact the high school data SUPPORT our assertions, and we provide them for the interested reader in Note 3).
- Regarding the interpretational "worries" of Linn et al., after noting some cautions, Linn and his colleagues provided the following summary for their analyses, "The evidence reviewed provides strong support for the conclusion that norms obtained for grades 1-8 during the late 1970's or early 1980's are easier on most tests than more recent norms." So, student achievement is UP on commercial tests, and that is exactly what we concluded.

To summarize then, when one actually looks at the additional evidence alluded to by Stedman, one discovers that he has misrepresented some of it and that none of it generates insights that would have caused one to question the conclusions we stated in TMC--and with which Stedman states agreement. Truly, when it comes to challenging our statements about the myth of achievement decline, Stedman has labored mightily and brought forth a mouse.

## CREATING MOLEHILLS, PART TWO--THE MYTH THAT AMERICAN SCHOOLS ALWAYS FAIL IN COMPARATIVE STUDIES

Stedman also accuses us of making a fourth "sweeping claim"--that "U. S. students 'stack up very well' in international assessments (p. 63). This assertion is largely correct, although some context should be provided so that readers will understand what we did and did not mean when making this claim. In our analyses of the issues involved in comparative studies of student achievement, we made five general points:

1. Few of those studies have yet focused on the unique values and strengths of American education.
2. Many of the studies' results have obviously been affected by sampling biases and inconsistent methods for gathering data.
3. Many, perhaps most, of the studies' results were generated by differences in curricula--in opportunities to learn--in the countries studied.
4. Aggregate results for American schools are misleading because of the huge range of school quality in this country--ranging from marvelous to terrible.
5. The press has managed to ignore most comparative studies in which the United States has done well. (p. 63)

Of these general points, the first and third are particularly crucial. By comparison, the United States operates an education system that has many unique features which reflect the values of our nation. Americans value a broad education, and this means that they offer more curricular options in their schools and colleges and lay less stress on the early mastery of core subjects than do most other industrialized nations. They also value creativity, initiative, and independence of thought in students, so they (sometimes, though not often enough) support curricula and classroom practices that encourage these traits rather than conformity to arbitrary standards. Our country also seeks to serve the needs of a huge range of students--including those from many different ethnic groups and those with both talents and handicaps--and this places unique burdens on our public schools. Americans also believe that education should provide equal opportunities for all, and as a result we build a unique set of second-chance opportunities into our school systems. And because we value higher education strongly, we enroll a lot more of our young people into colleges and universities, and our graduation rates are the highest in the world.

Because of these reasons, and because most comparative studies to date have assessed only the achievements of younger students in core subjects, they have, in effect, managed to AVOID most of the true strengths of American education. Commenting on this situation, we wrote in TMC: "If Americans are truly interested in learning how their schools stack up comparatively, they should insist that at least some comparative studies focus on the values that AMERICANS hold for their children and the unique strengths of AMERICAN schools.... [To date] none of the studies seems yet to have investigated breadth of student interests or knowledge; none has yet examined student creativity, initiative, social responsibility, or independence of thought; and few have studied knowledge among undergraduates or young people who have completed their educations. In fact, comparative studies to date seem almost to have deliberately avoided looking at the strengths of American schools!" (p. 53). Given this biased focus, it is actually quite surprising that our country has done as well as it has in comparative studies of achievement, and it was with these and related thoughts in mind that we wrote, "The myth that American schools fail badly by comparison with schools in other industrialized countries is also not supported by the evidence. Instead, when we analyze that evidence responsibly and think carefully about its implications, we discover that American schools stack up very well" p. 63).

In his critique of us Stedman AGAIN begins by stating his general agreement with our position. He writes, "U. S. performance in the international arena is not as dismal as school critics have asserted." (If needed, additional confirmation of this point, on which Stedman and we agree, may be found in the recent thoughtful review of comparative evidence by Gerald Bracey, 1996). So once again, where's the beef?

Stedman seems not to have been concerned about the issues we raised in our first, second, or fifth general points summarized above; indeed, he ignores them completely and as a result again misrepresents the thrust of much of what we wrote. (To illustrate, he asserts that we either wrote or implied that American performance in comparative studies is generally "glowing." We neither wrote nor implied such a claim.) He does, however, take issue with our third and fourth points, again citing his own published studies, claiming that the latter made substantive points that would contradict some of our conclusions. We turn now to these latter issues.

For one, Stedman asserts that American students "have done well in reading and elementary school science, middling to poor in geography and secondary school science, and last or near-last in mathematics." Although we were familiar with some of these apparent effects when we wrote TMC, we decided that validity problems in the comparative research literature were so great that stating such detailed conclusions was not justified at present, nor did we include them in our book. So here Stedman is complaining about what we failed to assert. Moreover, we are far from the only scholars to have noted serious validity problems in comparative studies of achievement. A Japanese teacher of mathematics has recently discussed the serious difficulties of trying to equate samples of American and Japanese students and of the absurd results that can be generated by studies based on badly flawed samples (see Ishizaka, 1993). He questions Japanese superiority in mathematics and is amazed that Americans believe the results of such flawed studies. But who is this teacher? Why should we put any credence in his remarks? Kazuoko Ishizaka is his name, and he is Chief of the Curriculum Research Division of the National Institute for Educational Research in Japan (Note 4). Ishizaka also notes the errors inherent in the oft cited work of Stevenson and Stigler (1992), whom Stedman unwisely cites to support one of his stranger assertions about the supposed strengths of Japanese education.

For a second, Stedman characterizes our conclusion about opportunity-to-learn as a "red herring" and quarrels with our presentation of evidence that was originally generated by Ian Westbury (1992) from the Second IEA Study of Mathematics Achievement. In this presentation Westbury (and we) pointed out that the typical Japanese 13-year-old has taken algebra whereas the equivalent American student has not, thus aggregated mathematics scores for students of this age show Americans to be at a disadvantage; but when the American data are disaggregated to display achievements for students who have and have not taken algebra, the achievements of the former look quite similar to those of Japanese students. Surprise! Somehow Stedman takes this simple demonstration of the effects of differences in curricula and opportunity-to-learn and converts it into a series of assertions that we did not make in TMC and do not believe. To repeat our major point: Education systems in various countries offer sharply different curricula, differing sequences of courses, and differing opportunities to learn for students at a given age. These differences generate many of the so-called "findings" of comparative studies of achievement, and nothing that Stedman writes contradicts this general point.

For a third, Stedman misrepresents our general point about variability among schools in achievement generated by the enormous differences in levels of funding for schools in our country--an effect that should be less prevalent in most other countries where schools are funded more equally. Stedman asserts that we had argued that overall variability in achievement among students should be greater in our country, but we did not argue for such an effect.

For a fourth, Stedman objects to our graphic presentation of data from comparisons of NAEP and IAEP scores that were originally generated by NCES in 1993. The point we made in presenting those data was that they reveal HUGE differences in average achievement among the American states, and that those differences are comparable in size to differences among nations reported in comparative studies, with the achievements of the "top" American states looking rather like those of our "top" overseas competitors and the "bottom" American states looking like underdeveloped countries. To illustrate, average scores for Iowa, North Dakota, and Minnesota are right up there with the top performing Asian nations of Taiwan and Korea;

in contrast, Alabama, Louisiana, and Mississippi score right down there with the lowest performing nation, Jordan. To talk about an "average" score for our nation as a whole may therefore be misleading. Stedman doesn't like the implication of this conclusion, so he quarrels with details of the data generated by NCES (which we reported), but none of his quarrels vitiates the general point we made.

Finally, Stedman misinterprets arguments about the evil effects of poverty and prejudice on student achievements in America that we made repeatedly in TMC. He writes, "although racism and social inequality have taken a severe toll on many of our students' academic development, this does not explain the poor general performance of U. S. students... [and] even our top half have not kept pace internationally in math and science." Apart from the fact such statements utterly ignore the fact that poverty and racism are much greater problems in our country than in most comparable nations, why on earth would racist and social-inequality processes NOT depress the general, aggregate achievement scores of American students or the achievements of "the top half"? The mind boggles.

To summarize: In Stedman's assault on our review of comparative studies of achievement he chooses to ignore and in part to misrepresent what we had written, and again the substantive points he makes do not contradict those we actually wrote in TMC. Thus, as before, what Stedman writes represents a good deal of sound and fury but signifies very little. He has once again made molehills out of molehills.

## LIKELY MOTIVATIONS

We cannot know all of the reasons why Stedman would choose to write such an unfortunate diatribe--one clearly at odds with the many embarrassingly flattering reviews that the TMC has received. Some of the few who have so far criticized us had actually helped to create The Manufactured Crisis and presumably resent being found out and publicly scolded. Others apparently have bought into major myths we exposed in our book or derived and promoted inappropriate ideas for the "reform" of our schools, and must now defend their untenable positions. And some may possibly be miffed because we did not chose to cite works of theirs that they considered relevant to the arguments of TMC. However, it seems quite likely that at least a portion of Stedman's dyspepsia reflects yet another motivation. This becomes clear in the latter part of Stedman's "review" when he states that American school achievements are 'not good enough' and that the two of us should be chastised because we did not express this idea in TMC. He writes, "although achievement trends, for the most part, have been stable, academic and general knowledge have been at low levels for decades." And this leads him to claim that--in supposed contradiction to what we had written--"the achievement crisis is real."

This stance is a remarkably familiar one, of course. Indeed, school bashing has been a popular indoor sport in America for years, and in Chapter Four of TMC we offered numerous examples of such sour judgments about our country's schools dating back over much of the century. In addition, this critical stance adopts safe territory because the standards against which America's schools are to be judged and found wanting are arbitrary and can be made up as one goes along. And for this reason, as prominent neoconservatives have recently begun to discover that the myths of The Manufactured Crisis cannot be supported with evidence, their enthusiasm for this stance has blossomed.

Those who adopt this stance today tend to bolster it with three arguments. Some suggest that American schools have 'always' been weak achievers, and the fact that their achievements haven't risen recently should not be taken as a vote of confidence. Others--enthusiasts for standardized testing--delight in pointing out that 'too many' students cannot 'pass' those tests at a given level or correctly answer selected items from those tests. And still others claim that although present standards were all very well for the past, they are clearly inadequate for the demands of the future (which somehow are rarely explained). In his so-called "review" Stedman advances the first two of these arguments but, somehow, not the third.

Regardless of the arguments advanced, this stance reflects a value judgment, not evidence. Stedman is at least partly right, of course, in his suspicion that we do not share his values. We find it ludicrous that anyone should claim that "academic and general knowledge have been at low levels for decades" in this country. If this were actually true, how on earth did our nation ever manage to win World War II, send astronauts to the moon, create a plethora of new pharmaceuticals, and invent the transistor and virtually all the computer technology now used world wide? For that matter, how did we achieve the world's highest rate of industrial productivity, and establish ourselves as this century's dominant super-power? "Low levels" of academic and general knowledge? What nonsense!

In addition, as we made abundantly clear in TMC, we believe that America's long-suffering educators and hard-working students are more often the victims than the perpetrators of our country's serious and escalating social problems. We cannot believe that useful strategies for solving the problems of American education are likely to be promoted by unfairly scapegoating these deserving people.

On the other hand, Stedman seems to share at least some of our values. Toward the end of his missive, he writes: "To succeed in our most troubled communities, we will need to overhaul school financing systems and break down powerful, entrenched bureaucracies. But school reform is no substitute for job creation, income re-distribution, and political empowerment. We must make our educational efforts part of a broader social and political agenda, one that promotes full employment, community revitalization, and civic participation."

Such thoughts certainly parallel those we expressed in our book. Too bad that Stedman did not bother to ponder the implications of these latter ideas for understanding the enormous accomplishments of American educators who have persevered, indeed have often succeeded, in the face of escalating social problems that are FAR worse in our country than in other industrialized nations.

But regardless of whether Stedman did or did not agree with all the values we expressed in TMC, he should NOT have allowed such disagreements to generate the lacunae, misrepresentations, and trivialities that characterize his supposed "review" of our book. Indeed, one of the hallmarks of good scholarship is that it is both honest and careful in its portrayal of the works of others, even those works with which one disagrees. Either Lawrence Stedman is unfamiliar with the admirable standards expressed by Edwin Boring, or he chose to ignore them completely when writing his unfortunate review.

## A NOTE OF THANKS

We have both written books before, but this is the first time either of us has authored a work that is controversial. We have been truly startled by some of the distorted portrayals and outright lies that have surfaced in so-called reviews of TMC appearing in major media sources, but most of those sources do not provide opportunities for authors to correct such mischiefs. Thus, in closing, we would like to thank Gene Glass and the editorial board of Education Policy Analysis Archives for this opportunity to reply to Lawrence Stedman's disingenuous portrayal of THE MANUFACTURED CRISIS.

## NOTES

1. The SAT decline began in the 1960s. Left out of most arguments about the causes of the decline is the fact that a powerful new medium of education and entertainment came into play in the 1950s. Television viewing has consequences for cognition and effects on school performance. Because television entered the daily lives of children on a regular basis in the early 1950s, the first of the TV-raised generations to graduate from high school were the classes leaving the public schools in the early to mid-1960s. Coincidence? Probably not. The

work of Keith Stanovitch (1993) is relevant here. In a clever series of studies he shows that there is a high correlation between exposure to print and many kinds of performances on paper and pencil tests of general verbal information. If exposure to print went down in the 1950-1965 time period, then a reduction in verbal aptitude test scores would be expected. That is exactly what happened. And if the exposure-to-television hypothesis has any predictive power, then the verbal aptitude score decline should be greater than the decline in mathematics aptitude score. And that happened too.

Whether this sudden emergence of television in the lives of America's students did or did not have a depressing effect on average SAT scores will never be known. But it is clear that during this period the primary medium of recreation and instruction changed, and the SAT--originally calibrated in 1941--did not. The SAT is NOT a test of the ability to decode rapidly changing audio-visual information, though the cultivation of this aptitude has been required since the 1950s. The bottom line is this: two things changed in the 1960s, the medium through which students were acquiring most of their knowledge and the composition of the population electing to take the SAT. It seems more likely that the notorious "decline" reflected these two factors rather than any supposed drop in school quality.

2. Of the 24 scores (grades 1-12 in reading and in mathematics) for the median-level test-taker, the SRA tests show the following gains and declines from one norming to another: reading--up in four grades, down in eight grades, net loss 1.3 percentiles; mathematics--up in six grades, down in four grades, no change in two grades, net gain 1.5 percentile ranks. The average for all grades and both subjects on the SRA is a net gain of .2 percentile ranks per year for the median-level test-taker from one norming to another. On the SRA tests, then, what one sees is a tiny gain here and there, and a tiny loss here and there. But most important is that there is no discernible trend here at all. What on earth would readers have gained had we displayed these data in TMC?.

3. The estimated yearly change in percentile rank for the median test taker on the reading part of the California Achievement Test (CAT), from one renorming to the next, for grades 9-12, is: +2.1, +1.1, +.6, and +.1. Thus, in this case, every score reflects a gain. In Mathematics the comparable data are +2.0, +1.1, +.7, and +.3. Again, each year a gain is evident. And if we had included the Stanford Achievement Tests (SAT), we would have reported that yearly gain scores for grades 9-12, between one renorming and the next, were: for reading, + .8, 0.0, +1.0, +.8; and for mathematics, +1.0, +1.0, +1.0, +1.2. Which means that seven of the eight high school test scores were up, one was unchanged, and none showed a decrease. Thus we could have ENHANCED our claim about rising test scores for commercial tests had we included high school data on the CAT and the Stanford!.

The MAT reading tests generated mixed data for these four grades: scores were up in two grades, but scores were down in two others. The NET score in reading, however, was up, and ALL four high school grades provided evidence of increased scores in mathematics. So even had we included MAT high school data, our conclusion would not have been challenged. In sum, Stedman's claim that much was lost when we chose not to provide results from the high school level is false.

4. With some minimal editing to make his English clearer, Mr. Ishizaka said:

> Based on the entrance examinations, students [in Japan] can choose one of the high schools of [a] large attendance area. So naturally the high schools are ranked according to their academic abilities. In the top ranking high school of the prefecture (state) where I taught, the average score of the newly entered students would ordinarily be 98 or even 99%. Almost all students got full marks. In my school, I taught the part-time students who work in the daytime and study in the evening. The average score of those students is 2.1 [percent], just a little less than the average of all schools. The average when I participated in that test was just 3 [percent].

> In the Second International Mathematics Study [SIMS], Population B of Japanese students got extremely high scores. So many people believe that Japanese high school students do very well in mathematics. I have been teaching mathematics for ten years and I know how well they do. Their average on for the intended curriculum was just around 5 [percent] or less when I was a teacher of mathematics. That means that the majority of the Japanese high school students do not attain what is intended by the government. If you look [at] the Japanese textbook it contains lots of materials but it does not mean that the students attain all those materials. (p. 4- 5)
>
> [When] we pick...certain samples of students it frequently happens something like this....Japanese attainment trends of high school students...[are] something like the letter "U" shape. They are either doing extremely well or extremely bad. I told you when I make a test, the average score was less than 5 points. Five points when the full score is 100. But in some of the best schools the average score is 98 or 99%. High schools of Japan were ranked according to their academic ability, and students trying to enter science and engineering fields ordinarily attend top level schools. In addition, Japanese society is [strong on] academic credentials. What school he or she is coming from is very important. Therefore up to the time when they enter colleges and universities they study extremely hard. They study more than 2000 different kinds of test problems and remember how to answer those items. I myself had the experience of studying for the entrance examination. When we look [at the SIMS tests] the answer is choosing from among five choices. If we are practicing every day for the entrance examination, we know very quickly what would be the correct answer. If it is a written test, it would be a little different. Anyway, Japanese Population B samples [of SIMS] were chosen from these upper extremes. I am not a specialist of international comparisons. [But] I know what the high schools attainment trend [really] is. (pp. 6-7)

Mr. Ishizaka also notes that Dr. Merry I. White, a leading Japanologist has written something like this "The curriculum--the courses taken and the material covered--is so rich that a high school diploma in Japan can be said to be the equivalent of a college degree in the U. S." Mr. Ishizaka thinks that Dr. White has lost her mind. And Mr. Ishizaka also noted that the U. S. Department of Education, in one of its pamphlets titled AMERICA 2000 COMMUNITIES: GETTING STARTED quoted Harold Stevenson. Stevenson has made headlines many times claiming that in his comparison of fifth grade mathematics classes "The average score of the lowest Japanese classroom is higher than the highest American classroom average for arithmetic." (p. 13). Mr. Ishizaka simply thinks we are foolish to believe this. And he might have some relevant background for commentary on this issue since he not only taught in Japan and is a member of the Ministry, but he has had personal experience with U. S. schools. His own children attended Illinois public schools and found them to be great!

## References

Bracey, G. (1996). International comparisons and the condition of American education. *Educational Researcher*, 25(1), 5-11.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: Public, private, and Catholic schools compared.* New York: Basic Books.

Ishizaka, K. (1993). Japanese education--the myths and realities. Paper presented at the meetings of the Canadian Teachers Federation, Ottawa, Ontario, Spring, 1993.

Linn, R. L., Graue, M. E., and Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice*, 10(Fall), 5-14.

National Center for Education Statistics (1991). The condition of education, 1991.

Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement (NCES 91-637).

National Center for Education Statistics (1994). The condition of education, 1994. Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement (NCES 94-149).

Stedman, L. (1996) The achievement crisis is real: A review of *The Manufactured Crisis. Education Policy Analysis Archives,* 4(1).

Stanovitch, K. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. In H. W. Reese ( Ed.), *Advances in child development,* vol. 24 (pp. 133-180). San Diego, CA: Academic Press.

Stevenson, H. W., and Stigler, J. W. (1992). *The learning gap.* New York: Summit Books.

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? *Educational Researcher,* 21(5), 18-24.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Mark E. Fetler
*fetlerctc.aol.com*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 4** **March 15, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

# Standard Errors in Educational Assessment: A Policy Analysis Perspective

**Greg Camilli**
**Rutgers University**

*camilli@zodiac.rutgers.edu*

**Abstract:** In many educational settings, educational gains are measured and evaluated rather than absolute levels of achievement. Gains might be estimated for individual students,teachers, schools, districts, and so forth. In some educational programs, schools are required to make "statistically significant" progress over the course of one school year. This would typically require and estimate of the standard error (SE for short) of the gain, which is a number representing the precision of the gain similar to the "margin of error" in polls. Because SEs can be used to define educational targets, it is important to understand precisely what a standard error is -- and this requires going beyond the simple textbook definition. Statistical methods are tools for understanding social processes, but there is no necessary connection between a statistical method and an empirical outcome. A policy analyst must ask how closely features of the statistical theory correspond to aspects of the measured outcomes for a given purpose. For example, how much does it matter if the assumption of random sampling is violated in certain ways? Can one assume that the children or educators at a particular school during a given year constitute a random sample of some population that is perhaps spread across time, space, as well as cultural and institutional dimensions?

## INTRODUCTION

In many educational settings, educational attainment is measured and evaluated. The units on which these measurments are taken might be students, but also teachers or school districts. The TVAAS program (Tennessee Value-Added Assessment System) is a statewide assessment program that incorporates features such as these. Specifically, it is a "gain-oriented" statistical tool for collecting and analyzing student achievement test score data; that is, gains are the focus rather than absolute levels of achievement. The information provided by the statistical model is used in the evaluation of teachers, schools, districts, and the like.

It is not the purpose of this document to evaluate the TVAAS program itself. However, during discussion of this program on EDPOLYAN (Educational Policy Analysis Forum -- an electronic forum in which discussion is conducted on the Internet), the issue of standard errors (or SEs) arose (SEs are computed with the TVAA multi-level model). In particular, some schools are required to make "statistically significant" progress which entails a 1.5 to 2.0 SE gain over the course of one school year. Because SEs are used to define educational targets, it is important to understand precisely what a standard error is -- and this requires going beyond the simple textbook definition.

In most introductory statistics courses, the terms "population," "random sample," and "sampling distribution" are taught. No two samples give the same result, e.g., the average height of a sample will always differ to a greater or lesser extent across random samples. This is why polls always append the "margin of error" to reported percentages. For example, the polls report that 51% of respondents would have voted for Bill Clinton with a margin of error of plus or minus 3%. (The margin of error is the product of two components: the standard error (SE), and the critical value (CV). The former represents how much a result may vary from sample to sample while the latter is used in conjunction with the former to place a band of confidence around the obtained result. For the given example, 3%=SE*CV; and the band of confidence extends from 51%- 3% (48%) to 51%+3% (54%). It is common to interpret the latter by saying that even though there might be variation from sample to sample, 95 out of 100 samples would give a result between 48% and 54%.]

Why should the standard error (SE) serve as a standard against which gains are evaluated? This question must be answered at both a technical and policy level:

- Technical. If there were 10 people in a room and you wanted to know their average income, you could ask all 10 people and calculate the mean. But now suppose that a town has 500,000 people in it. Obviously the same strategy is not going to work. Rather, you must economize by sampling a representative cross-section and calculate the mean of this group. This method doesn't guarantee an accurate result all the time, but it does well most of the time -- especially with larger samples. Thus, a sample result is not an exact answer to one's overriding question. From a statistical point of view, numbers are fuzzy rather than precise creatures; and a statistician's concern is to keep the amount of fuzziness to a minimum.

  By metaphor, an exact number is a pin prick (or puncture) whereas a fuzzy number is a bruise. Two pin pricks would be easy to discriminate visually, but if two bruises were large and overlapping, they might be difficult to distinguish. Now imagine the radius of a bruise as the standard error: as the radius decreases, it becomes clear as to whether there is one bruise or whether there are two. And as the radius decreases to near zero, the bruise becomes a pin prick. In short, the standard error is the statistician's criterion for the separability of two numbers, and two numbers are conventionally thought of as separable if they are at least 1.5 or 2 SEs apart. This is equivalent to requiring that the confidence bands around two numbers not overlap.

- Policy. Statistical methods are tools for understanding social processes. There is no necessary connection between a statistical process and an empirical outcome,so policy analysts must ask how closely features of the statistical theory correspond to aspects of the measured outcomes of an educational program. An important part of this analysis concerns how well sentences typical of the statistical theory support actions based on the separability criterion of 1.5 or 2.0 or some others numbers of SEs.

  For example, a typical sentence of classical statistical inference might read "A random samples is taken from a population." To which the analyst might respond "What is the population?" Furthermore, if the population can't be defined, one might conclude that it is not possible to determine whether the sample was indeed random. Thus, the language of the statistical theory might not satisfactorily explain the SE criterion, in which case more analysis is necessary to arive at a pragmatic understanding.

These issues are explored in the following discussion among members of EDPOLYAN.

The discussants are in alphabetical order Greg Camilli, Sherman Dorn, Gene Glass, Harvey Goldstein, Bill Hunter and Leslie McLean. Passages have been edited to focus on the issue of standard errors. The original postings contained more ancillary issues as well as parenthetical comments. However, the participants have reviewed the following text for accuracy and completeness. In addition, further summary comments were provided by Harvey and these are given at the end of the discussion section. Original messages were posted in late December, 1994, through January, 1995.

## EDPOLYAN DISCUSSION

*Goldstein*: I have come in on what I gather is the tail-end of a discussion of missing data in the analysis of TVAAS system data to produce estimates of school effects. Apologies therefore if the issue has been discussed already, and also because I'm from a different educational system but one where we have had quite a lot of debate about value added analysis using longitudinal data.
... in the UK the value added debate has been looking at problems with the sampling errors (standard errors) of value added gain scores...it turns out that these are typically so large that you cannot make any statistically significant comparisons between most of your schools...only those at opposite extremes of a ranking. Is this also the case in Tenessee? If so what do you do about it when reporting?

*Camilli*: I've been wondering what the standard errors mean. Usually, I have in mind that a sample is drawn from a population, and an effect (say gain score) is estimated from the sample data. The standard error then conveys how precise this estimate is (much like the "margin of error" that pollsters use). For TVAAS, what are the sample and population?

*McLean*: The purpose of this post is to focus on two aspects of the TVAAS that I feel have received too little attention: validity and standard errors. This is not to say that the political nature of any evaluation is not important or to take anything away from the discussion of formative vs. summative evaluation.

Harvey Goldstein stated on Dec. 19, 1994, "..it turns out that these [standard errors] are typically so large that you cannot make any statistically significant comparisons between most of your schools...only those at opposite extremes of a ranking. Is this also the case in Tennessee? If so, what do you do about it when reporting?"

Below are listed the mean gains for math with their standard errors for schools within one of the larger school systems in Tennessee. These means are three year averages and were calculated from the TVAAS mixed model process. This should give an idea of the sensitivity of the process.

| TYPE OF SCHOOL | GRADE | RANK | GAIN | MEAN STD.ERR. |
|---|---|---|---|---|
| INTERMEDIATE | | | | |
| | 3 | 1 | 71.6 | 4.9 |
| | 3 | 2 | 71.2 | 3.7 |
| | 3 | 3 | 67.0 | 2.6 |
| MIDDLE | | | | |
| | 6 | 1 | 22.6 | 1.0 |
| | 6 | 2 | 20.1 | 0.8 |
| | 6 | 3 | 15.6 | 1.0 |
| | 6 | 4 | 13.9 | 1.0 |
| | 6 | 5 | 13.1 | 1.2 |
| | 6 | 6 | 12.5 | 1.0 |
| | 6 | 7 | 11.1 | 0.9 |
| | 6 | 8 | 9.8 | 1.1 |
| | 6 | 9 | 9.3 | 1.3 |
| | 6 | 10 | 8.4 | 0.9 |

```
   ......................
8       12   13.5       6.6
8       13   13.2       1.7
8       14   11.0       0.9
   ......................
```

The problem for those of us who have calculated, pondered and puzzled over such results as these, in national and international assessments, is that the reported standard errors are unbelievable (impossibly small). We can't say they are wrong, of course, because we lack the details of the calculations, but Harvey Goldstein has analyzed at least as much data and written several books and taken the lead in multilevel modeling (sometimes called, by others, hierarchical linear modeling), and his informed and experienced "opinion" is not to be taken lightly. The standard errors remind me of those Richard Wolfe found faulty in the first International Assessment of Educational Progress--the fault being that the estimates of error failed to include all the components reasonable people agree should be included. Moreover, the Std. Errors above are clearly proportionate to the mean scores, not a desirable outcome. There must be at least one error (three lines from the bottom of those displayed above). I, too, will leave to later a comment on the statement below from TVAAS, except to say that whatever it is they do is not "certainly sufficient":

*Camilli*: I thought that some of you might want to take a look at some statistics regarding the metric of the scores that TVAAS uses. Below, I've given the mean, median and standard deviation of the IRT metric for fall reading comprehension as reported in the CTBS/4 Technical Bulletin 1 (1989).(I hope this isn't too far out of date.)

| Grade | mean | median | STD |
|---|---|---|---|
| 1 | 473 | 481 | 84.3 |
| 2 | 593 | 606 | 81.1 |
| 3 | 652 | 657 | 59.6 |
| 4 | 685 | 694 | 53.6 |
| 5 | 707 | 714 | 48.6 |
| 6 | 725 | 730 | 43.8 |
| 7 | 733 | 738 | 43.6 |
| 8 | 745 | 750 | 43.1 |
| 9 | 760 | 764 | 38.6 |
| 10 | 770 | 774 | 39.6 |
| 11 | 776 | 780 | 38.2 |
| 12 | 780 | 782 | 38.0 |

If you plot these data by grade, some interesting possibilities emerge. For example, one wonders why students below average gain as much as students above average. The explanation I see is that there is much less room for growth at higher grade levels, but this is a function of the scoring metric. A transformation of scale might lead to different results.

*Goldstein*: I see that Les McLean and one or two others have taken up my query about how well schools can be separated taking the estimated standard errors into account. I don't yet know how the standard errors have been calculated, but based upon a table Sandra Horn xxx sent me, I would say that the results (e.g. for grade 3 based upon a 3 year average in one of the larger school systems, are in line with our own results. What you do (roughly) is multiply each standard error by about 1.5, use this to place and interval (i.e. +-1.5 s.e.'s) about each gain estimate and judge whether two schools are significantly different at the 5% level by whether or not the intervals overlap. Most intervals do! BUT if you average over 3 years then you get smaller standard errors so fewer do.

*McLean*: My observation that the gains were proportional to the standard errors does NOT seem to be true--within grades. If you lump all grades together, the correlation is over 0.5, but within grades (the correct plot, IMHO) it is essentially zero. Grade six shows a substantial NEGATIVE correlation, but there are only 12 observations.

What are these standard errors anyway? In a separate post to me, Greg Camilli points out that if all students are tested, then the "sampling error" has to be zero. What we need to make sense of this is, as I have said already, a technical report. How are they (TVAAS) modelling the error in their multilevel models? What explanatory variables do they use? Do they include covariance terms? Is the "standard error" an estimate of measurement error? Just how much data are missing?

*Goldstein*: Re Les Mclean's message about standard errors. He quotes Camilli as stating that the standard errors given are 'sampling errors' and that if all students are tested then these are zero. I am confused! The usual standard errors quoted in this context are those relating to the accuracy of the estimated school effects where there is a conceptually infinite population of students of whom those measured (whether they are all those in the school at a particular time or not) are a random sample. If they are not this, then what are they?

*Hunter*: [In response to Goldstein's last question] I cannot say what they are precisely, but I am quite confident that those in any particular school at any particular time are NOT a random sample.

*Camilli*: Harvey Goldstein is wondering what I mean about standard errors. TVAAS probably doesn't test random samples of students, or samples at all, given what I know about the program. Given that it's not a random sample, one could always imagine that this were the case anyway (a counterfactual): imagination is required by all theories of statistical inference. However, without some sensible restrictions any set of numbers whatsoever can become a "random sample from a population." And once the "population" is in place, it can be of any size at least as great as the "sample." Now if it is imagined as infinite, then we can go on to imagine our "sample" as one realization of an infinite possible samples. Thus, we arrive at an estimate of "sampling error" called the "standard error."

Here's a fictitious dialogue between an Educator and a Statistician regarding this point:

> E: What is this standard error associated with a gain score?
>
> S: It's like this. Suppose you have a conceptually infinite population of 8th graders, and from this population you took an infinite number of random samples and computed the gain for each sample. You'd want these to all give about the same result; it's like witnesses at a trial corroborating a claim made by the defendant. Small standard errors are analogous to a high degree of corroboration; while high SE's indicate a lot of uncertainty.
>
> E: What if the tested group isn't a sample?
>
> S: Then you just imagine it's a sample.
>
> E: OK, I'll imagine it's a sample, but what's the population?
>
> S: Just imagine that the population is pretty much like the sample.
>
> E: So if I get a small SE, then I can be confident in the gain score because most of the samples from the population will have similar gains, because the population is pretty much like my sample?
>
> S: Basically.

> E: What if the standard error is large? Does this mean that I shouldn't be confident because most of the sample in the population that is similar to my sample will give substantially different estimates of gain?
>
> S: Well, you've got the basic idea.
>
> E: Okay, but just one more question. If the SE is large, doesn't it mean that the population isn't similar to my sample? If so, how can I imagine an infinite number of samples from that population?
>
> S: Look, SE's are really theoretical quantities. They're things that are defined by equations -- and the equations can be explained in different ways. Population/sample is the easiest way, but don't get bogged down. Most statisticians agree that they are useful and that small ones are better than large ones.
>
> E: OK, but just one more question. How small do they have to be to be good?
>
> S: That depends on your question. Suppose you want to test whether one teacher's gain is larger than another's. If the difference is one the order of 1.5 or 2.0 SEs, then you can have confidence in it.
>
> E: Why should I have confidence?
>
> S: Because the difference is large relative to the sampling error ... er, I mean, standard error.
>
> E: I see. Well, I have to go now. By the way, could you write something up that I could give to parents that explains this? Thanks.

*Goldstein*: Well, I enjoyed Greg Camilli's imaginary conversation, but of course the reality is that standard errors are not things statisticians invented to make life difficult. Most non-statisticians have little difficulty in understanding that if you only have a measurement on 1 student there ain't much to be said about the rest. The bigger the sample the more confident you become that what you have observed is a good guide to what you would get on repeated samples with also suitably large numbers...assuming of course that you adopted a sensible randomly based sampling strategy.

Now we come to the philosophical bit. Social statisticians are pretty much forced to adopt the notion of a 'superpopulation' when attempting to generalise the results of an analysis. If you want to be strict about things then the relationship you discovered between parental education and student achievement back in 1992 from a sample of 50 elementary schools in Florida can only give you information about the physically real population of Florida schools in 1992. Usually we are not interested just in such history, but in rather more general statements that pertain to schools now and in the future...we may be wrong of course and that is why we strive to replicate over time and place etc. BUT the point is that, getting back to value added estimates for a school, if we want to make a general statement about an institution we do have to make some kind of superpopulation assumption....what we happen to observe for the students we have studied is a reflection of what the school has done, and would have done, for a bunch of students, given their measured characteristics such as initial achievement. The more students we measure the more accurate we can be and that's why we need an estimate of uncertainty (standard error).

*Glass*: Greg answered with a hypothetical conversation between an educator and a statistician. I think Greg exposed some key problems with this notion of standard errors, and it is no more a problem with TVAAS than it is a problem with most applications of inferential statistics in education.

Harvey asks, in effect, what is wrong with regarding standard errors as being measures of the accuracy of samples as representations of "conceptually infinite populations" from which the samples might "conceivably have been drawn at random."

After more than thirty years of calculating, deriving, explaining and publishing "standard errors" and their ilk, I have come to the conclusion that I don't know what they mean and I doubt seriously that they mean anything like what they are protrayed as meaning.

Consider this: if the population to which inference is made is one that is conceptually like the sample, then the population is just the sample writ large and the "standard error" is much larger than it ought to be. If you show me 25 adolescent largely Anglo-Saxon boys who love sports and ask me the population from which they could conceivably have been sampled, I'll conceive of an "infinite" population of such boys. If no population has actually been sampled and all I know about the situation before me is the sample, then I will conceive of a population like the sample. This is surely the very opposite of inference and standard errors are surely beside the point.

Consider something even more troubling: I present you with a sample- - Florida, Alabama, Tennessee, South Carolina. N=4. I calculate the state high school graduation rates, average them and calculate a standard error. What is the population? States in the Southern U.S.? Fine; that's certainly conceivable, even if not "infinite." But suppose that someone else conceives of "States in the U.S." Well, that's conceivable too. But it is surely ridiculous to think that these four states can be used to infer to both of these conceivable populations with equal accuracy (standard errors). Or to make matters worse, suppose that I suddenly produce a fifth "state": Alberta. Now it raises the question whether the conceptual population is "geo-political units in North America"-- or the entire Western Hemisphere.

I can't imagine that there is much wisdom in attaching a number accurate to two decimal places when we can't even be certain whether it is referring to an "inference" to the Southern U.S., North America or the Western Hemisphere. Now, if you think I am playing with your head and will suggest a way out of this dilemma that rescues the business of statistical inference for us, let me assure you that I have no solution. In spite of the fact that I have written stat texts and made money off of this stuff for some 25 years, I can't see any salvation for 90% of what we do in inferential stats. If there is no ACTUAL probabilistic sampling (or randomization) of units from a defined population, then I can't see that standard errors (or t-tests or F-tests or any of the rest) make any sense.

Does any of this apply to TVAAS? Just this. If one is worried about "stability" (in any of the many senses in which the word could be interpreted) then why not simply compare teachers' scores across all years for which data are available. That would answer in very straightforward ways whether the ranking of teachers jumps around wildly for whatever reasons or is relatively steady. (I hasten to add that I don't approve of such things as ranking teachers with respect to their students' test scores.)

*Goldstein*: Gene Glass also takes me to task on standard errors and raises the interesting question of when a sample should be considered as having a reference population and when not. There is no general answer...it depends on what you want to do. As I said in my response to Greg, I cannot easily see how you can have empirical social science without assuming that the units (people, schools etc) you happen to have measured are representative (in the usual statistical sense) of a (yes) hypothetical population whose members exhibit relationships you want to estimate. Such populations must (I think) be hypothetical because they have to embrace the present and future as well as the past when the data were collected. The issue is therefore the general philosophical issue and not a statistical one - statisticians simply try to provide tools for making inferences about such populations.

*Camilli*: Harvey replied to my previous post with "The bigger the sample the more confident you become that what you have observed is a good guide to what you would get on repeated samples with also suitably large numbers...assuming of course that you adopted a sensible randomly based sampling strategy."

Bigger is better, I agree. Another issue is whether it is the correct standard error, and still another is whether the SE has a meaningful referent. If the sample consists of all kids in the system, how can imagining a larger group possibly create more information. If I want to understand the behavior of my three cars (I wish), how would it benefit me to imagine I had a fourth? This is not a statistical issue at all. "Population" has always been a heuristic device.

Generalizing beyond known populations is risky business, and requires more than statistical knowledge. This was the focus of the long and interesting dialogue between Lee Cronbach and Don Campbell. Standard errors have something to do with the precision of estimates. Perhaps they convey something about how well a model fits certain data. You might want to argue, on this basis, that the model is likey (or not) to generalize; but model fit at one instant does not logically imply model fit one second later. This, I think, is the difference between induction and inference.

The standard errors will apparently be used to measure whether statistically significant progress is being made by schools that fail to meet the standard (whatever that turns out to be), so it is important to be clear about what SEs mean. I find it fascinating that they are being used as policy tools with legal implications. In this regard, it is important to understand what drives the SEs. I'm guessing that missing data will add to SEs (it really would be helpful if the TVAAS staff would respond), and am sure that unit size will decrease SEs. Thus, standard errors for schools will typically be smaller for districts than for schools than for teachers than for students. As far as I can tell, only certain districts are required to make statistically significant progress; this may turn out to be a pretty easy criterion to satisfy.

*Goldstein*: When you try to enshrine complex technicalities in the law you certainly ask for trouble - especially, as would appear here, when those drafting the law have a rather meagre understanding of the technicalities. My interpretation of $49-1-601 is that it requires (say from one year to the next) that the differnce in value added scores for a school between two years is statistically significantly different from zero (at 5%?). If each years scores are on the same metric then this question can certainly be asked and one can even think of a suitable interpretation. The problem arises if we require this to be the case for all those schools below the mean (note that the legislation does not say STATISTICALLY SIGNIFICANTLY BELOW the mean.). If the schools are successful then the mean for all schools inevitably goes up!! and it isn't difficult to envisage a scenario where every school makes a real (even statistically significant) gain leaving the ranking of all schools the same! This raises the issue of the measurements used. Are these standardised each year on the Tennessee population? If so then not only is the ranking the same, so are the actual scores! All this needs some careful unpicking I would have thought and raises very serious issues for the interpretation of TVAA.

*McLean*: The discussion of standard errors has gotten so involved that a look at the Tennessee legislation should tell us where standard errors are needed and what interpretations reasonable people ought to be able to put on them. Below, the text from Sherman Dorn's post [who is quoting and paraphrasing from TVAAS statutes] and Les McLean's reponses are indicated by "-->".

> -->*Dorn*: The goal is for all school districts to have mean gain for each measurable academic subject within each grade greater than or equal to the gain of the national norms.
>
> -->*McLean*: How will anyone decide whether the mean gain is greater than or equal to the gain of the national norms? Publication of "standard errors" must mean that an error bound will be established around the national norms--perhaps 1.5 Times the median std. Error per grade--one "harvey", or 2.0 Std. Errors--one "dorn".

-->*Dorn*: If school districts do not have mean rates of gain equal to or greater than the national norms based upon the TCAP tests (or tests which measure academic performance which are deemed appropriate), each school district is expected to make statistically significant progress toward that goal.

-->*McLean*: ok, gang, the veil is lifted from our eyes--there is no such thing as "statistically significant progress" without standard errors and the assumption of samples from some population.

-->*Dorn*: schools or school districts which do not achieve the required rate of progress may be placed on probation as provided in $49-1-602. If national norms are not available then the levels of expected gain will be setupon the recommendation of the commissioner with the approval of the state board.

-->*McLean*: Yo, commish! I do not envy you your task.

-->*Dorn*: value added assessment means: (1) a statistical system for educational outcome assessment which uses measures of student learning to enable the estimation of teacher, school, and school district statistical distributions; and (2) the statistical system will use available and appropriat data as input to account for differences in prior student attainment, such that the impact which the teacher, school and school district have on the educational progress of students may be estimated on a student attainment constant basis.

-->*McLean*: I could write a rationale for a "statistical system" that did not need standard errors, given that they test all the students. It would contain careful, modern descriptive statistics that would gladden John Tukey's heart.

-->*Dorn*: On or before July 1, 1995, and annually thereafter data from the TCAP tests, or their future replacements, will be used (notice the 'will'-- the language is not just permissive here) to provide an estimate of the statistical distribution of teacher effects on the educational progress of students within school districts for grades three (3) through eight (8).

-->*McLean*: Here we are again--these gains are to be interpreted as "teacher effects". Peace, TVAAS, but I do not believe that anyone's models and techniques are yet good enough to isolate the teacher effect from all the other effects on standardized test scores in schools with all their complexity. Next to this concern--it is a concern about validity and is not vague or complex-- the definition and estimation of standard errors is too small a matter to take our time.

*Goldstein*: Les McLean's comments have inspired some more thoughts. In the simplest value added model, an outcome score is regressed on an input score so that generally each school will have a different regression line - perhaps with varying slopes but in the basic model with parallel slopes so that schools can then be ranked on the resulting regression intercepts. (The actual analysis is a bit more complex but this simple model captures the essence). We find, typically, that the variation among these intercepts is relatively small compared to the residual variation of student scores about the regression lines for each school (5% - 30% depending on which educational system you are studying). In addition, the regression itself will account for quite a lot of the variation in outcome...maybe as much as 50-60%.

This means that there is a substantial remaining variation (among students) unnacounted for and it is this residual variation which determines the standard error values. Thus, for example, if this residual variation was zero, we would exactly predict each schools (relative) mean and the standard error of that prediction would be zero. This would mean also that once we knew each student's input score (and anything else we were able to put into our regression model) and the school that student was in, we would have a perfect prediction of the student's outcome. Of course, we are nowhere near that situation and it is this uncertainty about the individual prediction that translates into uncertainty about the school mean (think of the mean roughly as the average of the student residuals about the regression line for each school). If you took another bunch of students with exactly the same set of intake scores you would NOT therefore expect to get the same set of outcome scores - this is what the uncertainty implies - nor the same mean for the school. In the absence of being able to predict with certainty we have to postulate some underlying value for each school's mean (otherwise we are pretty well lost) which we can think of as the limit of a series of conceptual allocations of students to the school. Thus an estimate of uncertainty, conventionally supplied by calculating the appropriate standard error,is important if you want to make any inference about whether the underlying means (that is, the population means) are different and, more importantly, to set limits (confidence intervals) around the estimated difference for any two schools or around the difference between a school's estimate and some national norm. Hence my original remark some time ago that when you did just that you found that most institutions could not statistically be separated, and I suspect also for TVAAS that very many cannot statistically be separated from a National norm, whether they are actually above or below it. It would be good to hear from the TVAA people on this issue.

*Camilli*: Harvey continues the standard error saga, and I want to reiterate: if you had all the students in the school there wouldn't be any uncertainty at all; you'd know the mean. I think we need a "superpopulation" to get us out of this predicament. Harvey said "If you took another bunch of students with exactly the same set of intake scores you would NOT therefore expect to get the same set of outcome scores - this is what the uncertainty implies - nor the same mean for the school."

This bunch of students is from the superpopulation, no? They are students who might exist, but don't, who are substantially like the students in the sample. I'll say it again, Harvey, this is a heuristic. It simply doesn't convey any additional meaning regardless of how many times it is repeated. I think we're lost when we accept statistical inferences based on data that weren't observed, and moreover, do not exist conceptually. If "all the students in the school" doesn't really have that meaning, then we are playing a game with language.

If we can get away from the superpopulation for a moment, we can begin to analyze what drives the standard error. It certainly isn't sampling error; nonetheless, it is a quantity that exists in a real sense. As you've implied above, SEs have something to do with model fit. Thus, we should be interested in those things that cause models to fit more loosely to the data. District size is certainly one factor; but correlation of effects within the model will also inflate SEs. Effects like teachers within schools, teachers with school, schools with district might be some examples. As Gene implies, separating these effects may take some doing.

*McLean*: Harvey Goldstein's exposition on standard errors (17 Jan, "Standard Errors: yet again") may have been more than some wanted, but I found it instructive and thought-provoking. If you deleted without reading, reconsider--it gets at the heart of the matter of TVAAS. While still wanting to retain the concept of the sample from some (unspecified) population, Harvey's main lesson for us was to highlight the crucial role of the model adopted by the statistician in estimating scores--gain scores, in the case of TVAAS. A model is a formula that the statistician considers a reasonable try at relating the desired quantity, the 'gain' in achievement (not directly measurable because of nuisances such as social class and prior learning) to aspects of schooling, such as teacher competence.

Advised by statisticians with wide experience outside of education (and maybe in education--we have not been told), the policy-makers decide to give the statisticians their head and to accept their estimate of 'gain', knowing that the formula will be complex and the procedures well beyond the understanding of all but a very few. The statisticians make a persuasive case that their formula and their procedures will provide the policy-makers with an estimate of gain that will distinguish the bad teachers from the poor from the average from the good from the excellent. "National norms" are invoked, unspecified, but responsibility given to the Commissioner of Education to provide norms if the national government lets the side down.

All this tedious repetition is needed to give a context for Harvey Goldstein's description of standard errors. In essence (correction, Harvey, please, if needed) the errors are S&E, not SE--errors of Specification & Estimation, not of sampling. A 'specification' error is made when our model, our formula, does not accurately link the target (the gain) with the data (the item responses or scale scores plus proxies for prior learning and social class and the like). We ALWAYS make a specification error--the only question is how large. If we limit ourselves, as in the TVAAS, to linear models, and we try to estimate gains across big, complex societies such as states, the error can be huge--and there is not consensus how to estimate the size of the error. Here is a source of error.

Even though they do not sample students and schools, sampling cannot be avoided--people are absent, times of testing vary, the tests cannot possibly cover all the content (hence content sampling), items are omitted, test booklets get lost, some teachers do not cover the material on the test, ..., and so on and so on and so on. This is why we do not use a very simple formula:

Gain = (Avg. score end - Avg. score beginning)

After all, when we test everyone, and when the goal is to measure gains by THESE students THIS year in THESE places with THESE teachers, who needs an error term? With well-constructed tests, the measurement errors will cancel out when we calculate school and class means. Oh--there is measurement error in individual pupil scores, but we can report that (from the test publisher's manual) and besides, these scores don't count in the student's grade--the teacher does not get them in time, and even if they do they do not use them.

Ok, so I seem to have lost the tenuous thread of the argument--NOT SO! We have learned over the years that the simple formula is more likely to mislead than to lead--to distort our view of gain rather than to clarify it. Raw score comparison tables (called 'League Tables' in the UK, after the rankings of sports teams), however compelling they seem, are statistically invalid, immoral, racist, sexist and stupid. Apart from those few flaws, they are fine. But would Tennessee put up with such poor procedures? Not on your life--scaling, imputation, hierarchical linear models and prayer are brought into play. Here is another source of error.

All this talk of standard errors and models and politics keeps coming back to one key aspect: VALIDITY. Do those numbers represent gains in achievement? The formulas and procedures are complex enough that evidence is needed. Even if they do, how accurate are they--and I mean how much do they tell us about better learning, class-by-class, teacher- by-teacher; or has the TVAAS traded in science for voodoo? Without a better explanation, the use of these scores to label teachers as competent or incompetent seems a lot like sticking pins in dolls.

It is possible to validate the numbers--but it would take a lot of thinking, a lot of hard work and maybe 0.01 of the budget of TVAAS.

*Glass*: Harvey, and are these future batches of students "random" or "probabilistic" samples from that "conceptual" superpopulation? It seems highly doubtful. So what sense can possibly be made out of probability statements that surely assume random sampling? None that I can imagine.

I think Les had it right last night. The "errors" in these teacher measurement schemes are model specification errors and not sampling errors. And the important questions to ask about them are not "will they be different in some conceivable 'population'?" but "what do they contain: ability differences in students, effects from previous teachers, etc.?"

*Camilli*: Les, I think your distinction between SE and S&E is a clear and elegant statement. It is a must-read for anyone interested in how statistical models are likely to behave in policy contexts. I'd like to throw in two additional cents:

1. I think TVAAS is certain to encounter a related problem with its "linear metric." How is it, the press may ask, that gains are so much larger in the earlier than the later grades? Does this mean that students aren't learning very much in high school? Moreover, because the standard errors are likely to be different across districts, larger districts might have to achieve smaller gains to be consistent with the law. Does this imply different standards for different districts? (I recognize that larger districts have to pull up more kids to achieve a SE's worth of gain -- but I'm not sure this type of argument would wash since a SE may be only a baby step toward the national average.)
2. The "natural" sample that exists on any given day does, I suppose, give rise to a superpopulation of the sort that Harvey Goldstein writes of. However, this is not the population about which most people think of when evaluating gains since, as Bill Hunter points out, it is not a random sample from the school's student body.

*Hunter*: Per Camilli who wrote "The "natural" sample that exists on any given day does, I suppose, give rise to a superpopulation of the sort that Harvey Goldstein writes of. However, this is not the population about which most people think of when evaluating gains since, as Bill Hunter points out, it is not a random sample from the school's student body."

I need to clarify a bit. I think it is not the case that a sample of convenience "gives rise to" or "implies" a population of any sort (unless one chooses to regard the sample _as_ a population). As far as I can tell this thinking is exactly backwards--samples derive their meaning and existence from populations: I cannot see that the reverse order has any meaning at all. I also question the utility of Harvey G.'s conception of such samples as samples from a population in time. This _might_ make sense in a time/space of great stability, but I see little reason to believe that children four or five years from now will have experiences of the world (especially the world of information) that is comparable to children of today (or five years past). The kinds of changes that required revision and re-norming of intelligence tests every 15 or 20 years half a century ago now take place in five years or less--probably about the same time scale that would be required to conscientiously develop and renorm the test.

Moreover, I think it is not just that such a sample is not a random sample from some _specific_ population (as Greg suggests above), but that it is not a random sample of ANY population for two reasons: 1) the process of selection did not insure equal and independent likelihood of selection for all members of the population and, more importantly, 2) no population was specified (to which the above process was not applied).

*Goldstein*: Brief response to Greg. The point about imagining another bunch of students like the ones you used to compute the school mean is that this seems to me just what one always has to do. The information about the students whose data you analysed may be of historical interest, but for most people they really want to assume that, given no evidence to the contrary, if and when a fresh set of similar students passes through the school (as is happening by the time they get to read the report)they would expect a similar outcome. The superpopulation is not just a heuristic device it is a reality in th esense that further batches of students are samples from it. How else would you make sense of anything?

Now to Les' points: Specification error actually, I think, sits on top of what I mean by standard errors, the latter assume that the specified statistical model is a good description. This raises what I think is perhaps the more important issue. Are we using the right measures? have we adjusted for all th econfounding factors? Have we adjusted properly for measuring errors (unreliability). On this side of the Pond we have I believe won the intellectual (not the political - we are used to losing that one)argument against Les' RAW league tables and are beginning to make people aware of the limitations of value added ones. The standard error argument is only one point of reference but it is quite important because it does, I believe, point out the inherent scientific limits to any kinds of institutional comparison in terms of how finely ranked you can get. There is a kind of uncertainty principle operating; you can establish that there are institutional variations without being able to determine exactly which institutions are actually different from each other. That's perhaps difficult to live with but does seem to be a fact of life.

*McLean*: On January 18, Bill Sanders wrote (via Rick Garlikov--and along with many other topics):

> --> *Sanders*: To Leslie McLean, your plots of standard errors as calculated make no sense. Middle schools in the example school system we provided have more students than intermediate schools in almost every case. Thus, their standard errors tend to be lower. Middle schools also have smaller expected nominal gains. Therefore, your attempt to show a relationship over grades is nonsense.
>
> --> *McLean*: It was indeed the point I was making--that the plot (or correlation) over grades made no sense. That is why I argued that the within- grade correlations were the ones to look at--and that they were around 0.0. BTW, if means in a table are based on widely different Ns, you would do your readers a good turn to say so, don't you think? Your remark that "middle schools also have smaller expected nominal gains" is ambiguous and interesting. In what sense "expected"; in what sense "nominal"?

*Camilli*: About superpopulations: these are entities that don't exist, except in the imagination. Yet it is contended that it is a "reality in the sense that further batches of students are samples from it. How else would you make sense of anything?" A lot of people have sought to answer this question, among them Alan Birnbaum who paraphrased the likelihood principle as the "irrelevance of outcomes not actually observed." He went on the write of the "immediate and radical consequences for the everyday practice as well as the theory of informative inference." As for the superpopulation, it exists in one's mind as a vehicle for generalization. But generalization itself requires more worldly knowledge. For example, consider the standard error of statistic calculated from a poll during an election. You might say a population exists, but only for a limited amount of time. Experience with the rate of change in public sentiment (and the way the question is asked) is required for a valid generalization. Happily, however, we are in full agreement on the role of specification error, as masterfully articulated by Les.

## SUMMARY COMMENTS BY PARTICIPANTS

*Goldstein*: I am a bit confused by the TVAA requirement to make a gain of 1.5-2.0 STANDARD ERRORS. Shouldn't this refer to STANDARD DEVIATIONS? The standard error is a measure of the accuracy with which a statistic (e.g. mean gain score) is estimated. The standard deviation is a measure of population spread and is the appropriate unit to use.

*Camilli*: Sherman, a question has come my way from Harvey Goldstein. He asks whether STANDARD ERROR should be STANDARD DEVIATION?" It's my recollection that the law specifically states that SEs are to be used for assessing gain, not SDs. Could you send me the relevant section?

*Dorn*: Okay, here is the relevant section of the TN law, and the answer's "none" -- at least explicitly:

§49-1-601. (c) If school districts do not have mean rates of gain equal to or greater than the national norms based upon the TCAP tests (or tests which measure academic performance which are deemed appropriate), each school district is expected to make statistically significant progress toward that goal.

But statistically significant is a strange concept for TVAAS, since there is no random sampling -- it's supposedly everyone in the relevant universe. Does it mean statistically significant considering test-retest reliability? Does it mean statistically significant considering the norming population? Does it mean statistically significant considering a hypothetical "let's pretend this is a random sample" thought experiment? Yeesh. In point of fact, courts have not had a chance to even consider this, since probation is not a question until this fall, and the legislature has delayed individual teacher reports for an additional year, at least (Nashville TENNESSEAN, 31 May 1995). I find it amusing that a state court will decide what statistical significance is here.

*Goldstein*: The discussion has certainly been interesting and useful for me in forcing me to be explicit about a number of 'taken for granted' assumptions. There seems to me to be three separate issues being debated.

1. If we have a collection ( sample) of individuals on whom we make measurements, is there some sense in which we can and should regard these as members of a larger collection or population of individuals. Does this population have to exist in reality (i.e. it can be enumerated in principle) or can we think of a hypothetical 'superpopulation' and when might this be useful
2. If we accept that there is a population about which we may wish to say something (e.g. what is the mean gain score among ALL 6-7 year old boys), how can we obtain a RANDOM probability sample so that we can then apply the statistical techniques which require such samples in order to draw valid inferences?
3. Any member of a sample of human (social) beings simultaneously belongs to more than one recognisable population; thus a child belongs to a particular social background grouping and a neighbourhood of residence and a school etc. Which is the appropriate population for inference?

Let me tackle 1) first. There are clearly some real enumerable populations which we can sample and make statements about. Surveys of voting intentions are a case in point where we wish to say something about how the whole (voting) population thinks, based on a suitable (preferably random) sample. A great deal of statistical sampling theory exists to help us do just this. At the other extreme you have something like what has been called 'generalisability theory' in educational testing that chooses to regard a set of chosen test items as being 'sampled' from some (conceptually) infinite population of such items contained within a 'domain'. I personally have great difficulty with this concept since, as Gene Glass points out, what people seem to be doing here is to imagine the population as just a larger version of the items they happen to have (unless they really have sampled, for example words from a dictionary for a spelling test). This then tends to come down to a sleight of hand whereby you choose your own test items, declare that they allow you to make inferences about an undefined domain, and then use statistical procedures to describe how accurately you have been able to describe that domain. And don't believe them when they tell you they have rules for generating the items and the rules implicitly define the domain - it doesn't work!

There are, however, other cases where I simply don't see how you can make any substantial progress without the notion of a hypothetical population of which your sample is a realisation. In effect this is nothing more than saying that you want, on the basis of what you observe on one group of individuals, to make some statements about other, unobserved individuals. If you are doing ethnographic, case study research, you are interested in what you find for what it may tell you about other (similar?) cases. Likewise, if you observe a relationship between race and school achievement you are concerned to make a more general statement and set of speculations about the relationship as it may apply to other children It seems to me that without this there is no empirical social science possible. This is a philosophical not a statistical issue. If you can't make inferences about future individuals then all social science is just descriptive history. The notion of a superpopulation is simply a formalistaion which allows us to use the tools of statistical inference. It is the ONLY formalisation I know of which allows a satisfactory method of generalising from the observed to the unobserved.

This leads to the next issue, which is how one can conceive of drawing a random sample from such a population. Gene's example of the four US states as a sample is instructive. Suppose, instead of merely calculating the mean graduation rate across States you compared the probability of graduating between Florida and Tennessee. In 1994 you found a moderate difference. You might be happy to stop there and leave it at that. On the other hand as a social scientist you might want to contextualise the difference, noting that Florida and Tennessee had different social compositions and you wondered whether these might 'explain' the observed differences. You might go on to look at other factors, and soon you would be constructing quite sophisticated statistical 'models'. The point about these models is, in general, that they don't explain all the differences - there is residual variation between children in whether or not they graduate.. We might, in principle, be able to explain everything but in practice this is extremely rare, and Les McClean's discussion of specification errors is relevant here. So the unexplained variation is assumed to be random - a reflection of our ignorance if you like. It is at this point when we invoke the statistical assumption of random variation that we are forced to assume some kind of sampling (or exchangeability if you insist on being a Bayesian) from a population. Whether you wish to confine inferences to Florida and Tennessee or wished to make some tentative inference about the factors which 'explained' graduation variations in general, across time and space, is a matter of debate and presumably disagreement, but generalise we surely wish to do?

This gets into my third issue about the appropriate population of reference. In brief then, I am not arguing that we always need a superpoluation notion, which then leads on to the statistical apparatus of standard errors, etc., but I am saying that to make sense of school comparisons (as with State comparisons), adjusting for those factors extrinsic to schools (gain scores and much more than these of course, such as race and class and gender) the notion of a superpopulation is really indispensable for us to make any progress. Let me ask the question again which I don't think anyone answered: If you have two schools, each with 2 students following a particular course, who would stake their academic reputation on reporting a moderate difference in average (over 2 cases) gain score as a judgement that the schools were REALLY differentially effective? Or suppose there was only 1 student in each school? Of course this is extreme - I merely wish to pursue the logic of refusing to recognise a superpopulation to an absurdity.

*Camilli*: Harvey frames the discussion with three questions:

1. When should we think of sample as members of a superpopulation. Clearly, there is a way to draw a sample in which it makes sense to think of a sampling distrubution. Frequency does have meaning in this situation. But Harvey thinks that you can "make progress" by imagining a sampling distribution when the population is poorly defined and sampling isn't random, and "In effect this is nothing more than saying that you want, on the basis of what you observe on one group of individuals, to make some statements about other, unobserved individuals. If you are doing ethnographic, case study research, you are interested in what you find for what it may tell you about other (similar?) cases." But the logic here is circular: I will assume my sample is similar to other nonrandom samples, then I will assume that the results in these other samples (similar by assumption) will yield similar results. In short, I think generalization is possible, but classical frequency theory is a lazy metaphor. You can make inferences about the future, but don't think statistical theory provides a formal basis for this. Ian Hacking in The Emergence of Probability recounts how Hume demolished this notion in 1739 (see p. 181).
2. Models are proposed by scientists to account for variation, and few if any models fit perfectly. In this case, a measure of mistfit is a measure of ignorance, but whoa! How does one equate ignorance with random variation? It seems to me that this is an attempt to reify frequency theory. I agree that generalize is what we surely wish to do, but what is happening here is 1) a statistical theory is adopted which is a mathematical formalization, 2) a strict correspondence between the terms of the theory and real events is assumed, and 3) results and manipulations of the theory are presumed to have counterparts in the real world. That is, real world events are now assumed to follow statistical laws. (Marx is spinning like a top.) We will make progress when we can usefully distinguish descriptive theory from observed covariation.

Suppose one has two students from each of two schools with gain scores, yet knows nothing of how these students were encountered. Does one want to determine whether these students are representative of the schools to which they belong, or assume that they ARE representative of Population X? In the latter case, we are 100% certain that we have a valid sample; this is an easily recognized tautology. In the former case, we have more work to do. Generalization isn't impossible, but we must make an argument for doing so and defend its validity. The argument is based on evidence, completeness, and persuasiveness. None of these qualities is based in statistical theory.

## References

Hacking, Ian. (1975) *The Emergence of Probability.* Cambridge University Press.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

| | |
|---|---|
| John Covaleskie<br>*jcovales@nmu.edu* | Andrew Coulson<br>*andrewco@ix.netcom.com* |
| Alan Davis<br>*adavis@castle.cudenver.edu* | Mark E. Fetler<br>*fetlerctc.aol.com* |
| Thomas F. Green<br>*tfgreen@mailbox.syr.edu* | Alison I. Griffith<br>*agriffith@edu.yorku.ca* |
| Arlen Gullickson<br>*gullickson@gw.wmich.edu* | Ernest R. House<br>*ernie.house@colorado.edu* |
| Aimee Howley<br>*ess016@marshall.wvnet.edu* | Craig B. Howley<br>*u56e3@wvnvm.bitnet* |
| William Hunter<br>*hunter@acs.ucalgary.ca* | Richard M. Jaeger<br>*rmjaeger@iris.uncg.edu* |
| Benjamin Levin<br>*levin@ccu.umanitoba.ca* | Thomas Mauhs-Pugh<br>*thomas.mauhs-pugh@dartmouth.edu* |
| Dewayne Matthews<br>*dm@wiche.edu* | Mary P. McKeown<br>*iadmpm@asuvm.inre.asu.edu* |
| Les McLean<br>*lmclean@oise.on.ca* | Susan Bobbitt Nolen<br>*sunolen@u.washington.edu* |
| Anne L. Pemberton<br>*apembert@pen.k12.va.us* | Hugh G. Petrie<br>*prohugh@ubvms.cc.buffalo.edu* |
| Richard C. Richardson<br>*richard.richardson@asu.edu* | Anthony G. Rud Jr.<br>*rud@sage.cc.purdue.edu* |
| Dennis Sayers<br>*dmsayers@ucdavis.edu* | Jay Scribner<br>*jayscrib@tenet.edu* |
| Robert Stonehill<br>*rstonehi@inet.ed.gov* | Robert T. Stout<br>*stout@asu.edu* |

**Education Policy Analysis Archives**

**Volume 4 Number 5** **March 25, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# Opening up Jewish Education to Inspection: the Impact of the OFSTED Inspection System in England

**Judy Keiner**
**University of Reading**

*J.C.Keiner@reading.ac.uk*

**Abstract:** Although Jewish schools in England are generally deemed successful, internal communal surveys have highlighted concerns about their teaching of Jewish studies and modern Hebrew. The UK government in 1993 established detailed national criteria for four-yearly published inspections of all schools. This imposed the need to develop criteria for the evaluation of these specifically Jewish subjects, and both schools and foundation bodies have begun to respond through training and development activities. Analysis of the first published reports, shows evidence of mismatch between Jewish schools' aims for Jewish Studies and their practice. Common findings on modern Hebrew teaching indicate concerns about planning, methodology and assessment. The response of Jewish communal bodies is explored, showing an increasing focus and some rivalry towards servicing the inspection and development needs of Jewish schools. Jewish communal press reporting and parental response to inspection is considered.

## Historical background to the Jewish school system in England

England is always different. This statement is true for almost any aspect of education policy or provision you might care to analyse. The reason for that is largely to do with the particular history of English education, and the historical penchant of English policy and practice for combining evolutionary and incremental change. Not surprisingly, Jewish schools in England are different too. Since World War II, there has been a great rise in the number of Jewish primary schools established within the state system, which now includes twenty five state-aided primary and secondary schools, of which ch one new primary and one secondary school were established in the last three years (Note 1).There is a further substantial number of independent Jewish schools which do not receive any state aid, but which have tax-free charitable status. Of these schools, a small minority offer the similar combinations of secular and religious studies as their state-aided equivalent. The remaining schools are maintained by the most strictly orthodox, mainly separatist communities, including Chassidic communities, which in the UK number less than five per cent of the total Jewish community of around 275,000. The medium of instruction in many of these schools is Yiddish, and the courses of study are almost entirely centered on traditional sacred texts, with only a small proportion of time given to the teaching in English of English, mathematics and other secular subjects.

Five further new Jewish schools are in the advanced stages of planning, and plans to incorporate three formerly independent existing schools into the state aided system are also in their final stages. Whilst in the wider world, England is often assumed to be synonymous with the UK as a whole, the school system in Scotland is again different and autonomous. In Wales and Ireland, although very closely tied to the English school system, the school systems are under the auspices of the respective regional administrations. All Jewish schools in the UK are under the English administration apart from one primary school in Scotland.

The status of Jewish schools in England differs from other diaspora countries. In most countries, Jewish schools are private, receiving little or no state aid. But the history of mass provision for schooling in England began largely through the initiatives of Christian church foundation bodies setting up schools piecemeal, with dramatic rises in the number of schools in the wake of early nineteenth century industrialization. There was effectively an unevenly distributed but still nationwide network of church schools before 1850. The state began giving aid to these voluntarily established schools in the mid nineteenth century. As early as 1853 (Alderman (1989) p.16), England first gave the then very small number of Jewish schools state support, and then gradually absorbed them into the English state funded system (Note 2). This was achieved without any significant controversy (Note 3) as far as Jewish schools were concerned, since the state funding has always been solely for the secular subjects taught at the school, as well as a half of the cost of buildings. Such controversy as there was in the early years of the twentieth century, when the current state system of aiding voluntary schools was established, centered almost entirely on state subsidies to Roman Catholic schools, under the inflammatory banner of protests against "Rome on the rates".

From the end of World War I until the early 1960s, there were fewer than ten state aided Jewish schools in total, the vast majority of Jewish children attending secular state schools. That system offered much prized opportunities to enter elite educational institutions via competitive selection for prestigious state-aided day schools. This was the major route of social mobility and assimilation for the daughters and sons of Jewish immigrants, who were disproportionately successful in gaining places and scholarships.

## The rising popularity of Jewish schools since the 1960s

In the early 1960s, a combination of catalysts began to shift Jewish communal and parental priorities towards Jewish schools. There was an accelerating process of moving out from inner cities into outer suburbs, fueled by much wider availability of low-cost mortgages. Under the Labour administrations of that period, state selective schools were increasingly abolished or converted into fully comprehensive all-ability intake schools. There were the beginnings of media-fueled parental anxieties about ethnic conflicts and underachievement in schools as substantial communities from the "New Commonwealth" countries of the Caribbean and Indian sub-continent settled in three e UK, mainly in the inner cities and some of the outer London suburbs previously much favored by Jewish communities.

With the new growth in the popularity of Jewish schools at this time, the Zionist Federation Educational Trust (ZFET) emerged as the foundation body responsible for the largest number of Jewish schools (Note 4). By the early 1990s ZFET was the foundation body for four thousand children in their schools. The ZFET schools strongly promoted the teaching of Hebrew as a modern language, with a focus on Israel as great or even greater than that on the promotion of Judaism being their raison d'etre. The orthodox United Synagogue, the largest synagogal body in the UK, established a smaller number of schools in the London area, being responsible by the early 1990s for over two thousand four hundred pupils. Still other Jewish schools, particularly in the provinces, were independent organizations.

Jewish schools in the UK never followed any single agreed common religious education syllabus. The main Jewish voluntary organization responsible for religious education in the early post-war years was the London Board of Jewish Religious Education, founded in 1946, whose main responsibility was for organizing after-school and Sunday religious classes, at a time when there were relatively few Jewish state schools (Alderman (1989) p.105). The Board, which was closely connected with the United Synagogue, and was redesignated the United Synagogue Board of Religious Education in 1987 (Note 5), also formerly provided a syllabus for the teaching of religious education for Jewish children in local authority state schools in London, where the numbers were large enough to warrant the provision of classes by peripatetic teachers (Note 6). The influence of the Board syllabus was still detectable in the curricula of some Jewish primary schools when the National Curriculum (NC) was introduced in England and Wales at the end of the 1980s. The introduction of the National CurriculumThe National Curriculum has been one of the most far-reaching policy initiatives to affect education in England in the twentieth century. Prior to its introduction through the 1988 Education Act, the only legal curriculum requirements of schools were that they taught physical education and religious instruction. It also for the first time enshrined the principle of pupil entitlement, rather than opportunity, as the basis on which curriculum access was to be offered.

By the time of the National Curriculum, it is probably true to say that for secondary schools, the syllabuses for Jewish studies and Hebrew were effectively defined by the requirements of external school examinations. Few primary schools had religious education syllabuses which were other than a statement of the topics and reading skills set out in the old Board syllabus. In some primary schools, no written syllabus existed, and the curriculum was organized by reference to the Jewish calendar, with its associated agenda of weekly readings and festivals, and by whatever primers were used to teach reading of Hebrew for religious purposes. The National Curriculum is compulsory only in state and state aided schools, and so does not impinge directly on the independent schools. Nevertheless those Jewish independent schools which seek to combine secular and religious studies cannot avoid incorporating some of its requirements into their own curricula because of the requirements of entry to presitigious state schools and because public examinations assume a basic coverage of NC requirements.

## Recent dilemmas facing Jewish schools in England

Jewish state and state-aided schools in England have recently been in the headlines for very positive reasons. Jewish secondary schools in London and Liverpool have featured very prominently in the highest positions of the unofficial league tables, showing comparative results of examinations taken at 16 and 18, which the UK press has published over the last five years or so (Note 7). The schools are in very great demand by parents, with all but one or two schools, in areas of declining Jewish population, being substantially oversubscribed.In recent years, this apparently rosy picture has concealed a degree of communal and professional concern about the quality of Jewish religious and cultural education in the schools. In 1991 and 1993 respectively, the two major foundation bodies involved in state Jewish education, the United Synagogue and the Zionist Federation Educational Trust (ZFET) independently undertook reviews of Jewish education under their auspices (JEDT(1992); Hyman & Ohrenstein (1993) (Note 8). Both bodies came to similar conclusions about the problems, acknowledging a degree of lack of success in teaching both Jewish RE and both biblical and modern Hebrew, which are deemed essential for participation in prayer, and, in the case of the latter, for a relationship with the only Jewish state in the world, Israel. Both bodies acknowledged the need to remedy these shortcomings by developing major in-service programmes. The United Synagogue review additionally urged the setting up of a single educational agency for the entire Jewish community, which would in corporate the ZFET.

These initiatives marked the first effective move by Jewish foundation bodies into in-depth long-term strategy and policy making. It is interesting that their frames of reference were primarily those of corporate management; cost effectiveness and efficiency. There does exist within talmudic and other traditional religious sources a range of starting points which might be used for generating a policy analysis framework for Jewish education; these include references to the maximum size of classes, to what makes for educational success and failure, and to issues like competition and motivation.

Nowhere in either of the reviews was any reference made to these sources. It was not surprising that issues of teaching effectiveness were, along wit h those of curriculum management and resourcing, at the center of the short comings identified. Historically, the staffing of the teaching of Jewish religious education and Hebrew has been on a different basis from that of the staffing of the secular subjects in Jewish schools. Frequently, these two subjects have been taught by supernumerary specialist staff, whose sole role has been in either religious studies or Hebrew teaching. Their salaries have been paid by voluntary parental contributions, supplemented by subventions from the foundation bodies, which fund raise and, in the case of the United Synagogue, use a proportion of the substantial income gained from membership and burial ground fees. The staff often had no professional teaching qualifications recognised by the Department for Education and Employment (DEE). The religious studies staff in many cases obtained qualifications through private Jewish religious academies in Britain or in the USA or Israel, and the Hebrew staff often had Israeli teaching qualifications, albeit not qualifications for the teaching of Hebrew as a foreign language. The Hebrew staff have also frequently been short term placements sent from Israel, sometimes owing their placement to the fact that their spouses have been posted in England as representatives of Israeli government organizations. The organization and management of the schools has tended to reflect the different status of these staff. They have not usually held senior management responsibilities, and or taken responsibilities for pastoral work. Until relatively recently, they would frequently not have been involved in staff meetings or school based in service training days for the whole school.

## The implications of National Curriculum for Jewish education

With the passing of the 1988 Education Reform Act by the Conservative administration of Margaret Thatcher, the emergence of the National Curriculum came to pose particular challenges to Jewish schools. The 1988 Act maintained the careful delineation established in England and Wales of religious education, and particularly religious education in state-maintained schools run by voluntary religious organizations. The Act did not include religious education amongst its list of legally compulsory core and foundation subjects (Note 9), but recognised the continuing status of religious education as a pre-existing compulsory subject under the legislation of the 1944 Education Act. Thus, while legally binding specifications for what was to be taught at each stage of the curriculum were issued, in the form of printed folders, for each of the nine secular core and foundation subjects, the specification of the religious education curriculum remained as an evolutionary continuation of the pre-existing forms of local authority and voluntary foundation body control.

Day-to-day discourse in English schools and in the press about National Curriculum has almost invariably seen it as referring to the nine secular subjects, and not to religious education, which by reason of not having its own common national folder, has come to be seen as having less prestige and priority in the allocation of scarce resources for school development. Yet religious purposes were nevertheless central to the aims of the 1988 Education Reform Act, which in its opening clause refers to the requirement for "a balanced and broadly based curriculum which promotes the spiritual, moral, cultural, mental and physical development of pupils at the school and of society" (Great Britain (1988)).

While the Act itself explicitly excluded the specification of precise subject hourages, it did assign to each core or foundation subject notional pro portions of the curriculum time available in the school. The time thus allocated added up to some ninety percent of the curriculum, and a common complaint of head teachers and their staffs was that one hundred percent of curriculum time was not sufficient to deliver the legally required demands of the National Curriculum. Such pressures were the stronger on Jewish primary schools, where the time devoted to religious studies and to the teaching of modern Hebrew has usually been of the order of twenty to thirty percent of the school timetable.

### Dilemmas facing Jewish schools as a result of the National Curriculum

The response to this particular challenge of National Curriculum innovation varied amongst the Jewish schools, with the greatest pressures being on three e primary schools, which had not previously experienced the demands of externally defined curricular criteria. The inclusion of modern foreign languages amongst the foundation subjects of the curriculum potentially posed a major challenge to the teaching of Hebrew. The National Curriculum specification was based on current modern language teaching principles, requiring a substantial focus on developing pupils' ability to speak spontaneously in the target language. Hebrew teaching in Jewish schools has tended to focus strongly on reading and to some degree translation, since the reading of prayer books and the Hebrew bible are a central requirement of both Jewish religious education and Jewish practice. Moreover, the reading skills needed must encompass the classical Hebrew in which the bible and liturgy are written.

In practice, therefore, Hebrew teaching in Jewish schools has tended to be somewhat formal in nature, almost invariably based on highly structured graded readers and written exercises with controlled vocabulary. Because the Introduction of the National Curriculum was phased over several years, the specifications for modern languages were published only in 1991 and came into force in 1992. Modern languages were specified only for Key Stages 3 and 4 (ages 11-14) of the National Curriculum, and therefore the specifications appeared only to cover teaching in secondary schools. As previously stated, the impact on secondary schools was limited because their curricula have always been closely related to the demands of external examinations.

The Jewish schools responded to the pressures in a variety of ways, with three e responses in the primary schools ranging from a substantial extension of the length of the school day to, in the case of at least one primary school, a recognition that meeting the entire National Curriculum legal requireme nts was not compatible with its commitment to devoting twenty five percent of teaching time to Jewish studies and Hebrew, and that the legal requirement would not be fully met. The National Curriculum thus introduced the first stage of a modern national quality control system to English schools, in its precise specifications of curriculum requirements and assessment criteria, together with requirements to publish nationally moderated assessment results at specified points.

Because the NC was introduced over a phased period of five years, starting in 1989, the years 1989-94 saw almost all the development energies of schools focused on implementing one core or foundation subject after another. Each new subject implementation brought pressures on schools to review curricular provision and resources, with a legal requirement to produce a development plan setting out action programmes to bring any gaps in resources and provision into line. Finally, as a result of nationwide evidence of excessive workload resulting from the pressures described above, together with the growing organized teacher resistance to the implementation of the assessment system, the government instituted a major review which resulted in the slimming down of the NC to take up eighty rather than ninety percent of schools' curriculum time, to take effect from the 1995-96 academic year.

Already marginalized from the center of whole school initiatives for the reasons indicated above, the advent of the National Curriculum era served to widen the difference between the requirements and expectations of secular and of Jewish studies and Hebrew teachers in Jewish schools. The latter could see themselves as unencumbered by the straitjacket of National Curriculum legislation and its accompanying administrative work of assessment and record keeping. It might have been thought that Head Teachers and Governors, frequently feeling under great pressure with the volume of NC implementation, would feel it to be a positive benefit that two areas of the curriculum central to the raison d'etre of Jewish schools were not to be subjected to the same pressure of intensive review and adjustment which accompanied the coming into force of the secular subject regulations. However, as the NC process became embedded in the primary schools, Heads of Jewish schools could also see the opportunities given by the publication of national criteria and benchmarks for exercising a closer degree of quality control over Jewish studies and Hebrew than they had previously been able to do.

### The emerging incorporation of Jewish studies and Hebrew into national quality control initiatives

Two factors unforeseen at the time of the passing of the Education Reform Act came to shift the focus of curriculum priority in Jewish schools much more centrally onto Jewish studies and Hebrew. An initiative started from an internal Conservative administrative decision to review the role of Her Majesty's Inspectorate (HMI), which has always had a degree of autonomy from direct government control, in much the same way as the judiciary. It was seen at the time as possibly not sufficiently attuned to the educational vision of the Conservatives, and to some degree viewed with suspicion within the administration as being tainted with pro-teacher, pro-progressivist and anti-government perspectives, a bulwark of what the administration viewed as an entrenched educational establishment.

The review culminated in the replacement of HMI as the main agency of direct quality control inspection of schools with a new system of inspection by external teams of private contractors who would operate according to criteria set down by a new government agency for standards in education. A new Education Act, passed in 1992, established the new system of inspection, to take effect from 1993.

Secondly, the Secretary of State for Education who was in office at the time of this new legislation and until 1994, Mr John Patten, was not only a man of strong personal religious convictions but one who also advocated strengthening traditionalist religious education and Christian religious worship in schools as a bulwark against a supposed disintegration of societal values in Britain. During his period of office, religious education, previously all but neglected by his predecessors, and virtually ignored as part of the vast programme of National Curriculum training, became the subject of major new initiatives, including a requirement in the 1992 Education Act that religious education and worship in state schools other than those controlled by voluntary religious bodies, be in the main Christian.

Such initiatives can hardly have been implemented as the outcome of one politician's preoccupations, yet the initiatives were potentially explosive. For although religious education and religious worship had been compulsory under the terms of the 1944 Education Act, for many years very substantial numbers of schools had not carried out the obligation to hold a daily act of collective worship for all pupils. Indeed, the design of many modern secondary schools built over the last thirty years was such as to make it impossible to hold collective worship for the whole school; the largest assembly spaces in many of such schools are too small to seat the whole school simultaneously. Significant numbers of schools, particularly LEA schools in inner city areas, have not offered religious education on a regular timetabled basis, or where they have, it has frequently not followed the legally required Agreed Syllabus which each LEA had been required to establish for its schools under the terms of the 1944 Act.

### How the establishment of the new inspection system incorporated two historical traditions

The legislation implementing the new inspection system set out separate procedures for secular and religious education in schools controlled by religious foundations. Section 9 of the 1992 Education Act laid down procedures for the inspection of those aspects of any school covered by National Curriculum and other legislation, such as the Equal Opportunities Act and the Health and Safety Act. Section 13 of the 1992 Education Act laid down inspection procedures for the religious education which is wholly under the control of the governors and the foundation bodies of voluntary aided schools. This apparently strange separation of inspection procedures was the consequence of historical traditions of English state and religious schooling referred to above. The whole history of the status of voluntary aided schools has been rooted in an exclusion of state competence from any involvement in the specification or quality control of religious education in these schools. While such a distinction did not at first sight present any difficulties, there were profound contradictions built from the start into the 1992 legislation. For the 1988 Education Act itself carried in its first clause referred to above an obligation on all schools to provide for the spiritual, moral, social and cultural development of pupils.

These aspects of each school, broadly referred to as its ethos, were to be part of the Section 9 inspection. Yet for voluntary aided schools, the spiritual and moral, if not also the moral ethos of the school was surely derived substantially from its programme of religious education. The regulations allowed for the spiritual, moral, social and cultural aspects of the school to be inspected as part of the Section 13 inspection, if desired by the governors (Note 10). Nevertheless a further contradiction remained, for even in such cases, it was still to be the responsibility of the secular Section 9 inspection to report on whether the requirement for a daily act of worship for all pupils was being carried out, because of daily collective worship being part of the national statutory requirement for all schools. There was yet a further level of potential confusion and contradiction arising from the ambiguities of responsibility. Although the governors were given the option referred to above, confusion could arise because the arrangements for the two inspections could be made quite separately. It would not necessarily be clear to a Section 9 team whether arrangements for the Section 13 inspection to report on spiritual, moral, social and cultural aspects were being made, since there was no obligation to arrange the inspections to dovetail responsibilities.

The implications of the new inspection system, together with the new policy interest in promoting religious education only became fully clear from the academic year 1993-94 as the new government agency responsible for the organizations, the Office for Standards in Education (OFSTED), took shape under a Circular issued by DfEE defining its mode of operation (Great Britain-DfE (1993)) . One of the concerns expressed about the replacement of the former Her Majesty's Inspectorate, appointed by officially trained but independent contractors, was that schools would be able to choose contractors they deemed might be likely to write more favorable reports.

**The emerging inspection system and the choices open to governors of Jewish schools**

Circular 7/93 made clear that the system of contracted inspections would be handled by the OFSTED office itself, with OFSTED putting out tenders and a warding contracts for inspections of individual schools. Inspections were to be conducted by inspectors who had to follow a very detailed handbook (O FSTED (1993)), laying out criteria for the evaluation of every aspect of a school's performance. Each inspector would to have pass a rigorous training course designed to ensure their competence to apply the criteria and report according to procedures laid down in the handbook. However, this system was to apply only to Section 9 inspections. For the section 13 inspections of voluntary aided schools, it would be for the governors of each school to nominate the inspector or inspectors, and no criteria were specified for three e selection and competence of the inspectors, or of the inspection of the subjects.

It was thus to be open, for example, to the Governors of a Jewish school to choose to appoint, if they were minded to, the Prince of Wales, Ms Madonna Ciccione, a Governor's relative or a Jesuit priest to inspect their school's religious provision, and for that inspector to follow either the criteria laid down for the inspection of religious education in state schools or any supplied by the Governors, or none at all.

It was also left for the Governors of voluntary aided schools to choose whether or not they wanted the inspection of the religious side of the school's life to be inspected at the same time as the Section 9 inspection or not. Simultaneous inspection would be open to them only if they chose an inspector who had successfully completed the OFSTED training course. In this case it could happen only if the approved inspector contracted by OFSTED to lead the Section 9 inspection agreed that the Section 13 inspector could be part of the team. In such a case, the Section 13 inspector would also be able to have access to the full curriculum documentation which schools are required to provide as part of the Section 9 inspection. He or she could also take part in the team meetings which are an essential part of the inspection process in enabling inspectors to come to a consensus in judgements on the school.

There was another major and unforeseen implication of the OFSTED system for Jewish schools. This was the procedure adopted by OFSTED for inspecting subjects of the National Curriculum being taught for age groups other than those for which they were specified. It was that such subjects would be assessed in terms of the National Curriculum framework as published. Thus it emerged through processes of informal consultation with OFSTED that Hebrew taught in the Jewish primary schools for any amount of time longer than an hour a week would be assessed according to the specifications set out in NC Modern Languages. In the event, all modern Hebrew teaching in Jewish schools so far inspected has been reported on by OFSTED reports on this basis.

### The impact of the changes on Jewish education

How then have these changes impacted on Jewish education? There are two main sources of impact; firstly in the foundation bodies responsible for the schools, and in communal bodies closely involved in Jewish education. Secondly, there is the impact on the schools themselves, and on the wider Jewish community which they serve.

### From dilemmas to turf wars: the emerging response of the foundation bodies and communal organizations

As has already been noted above, the two major foundation bodies, the United Synagogue and the Zionist Federation Educational Trust(ZFET), had already mounted major reviews of Jewish education. In each case the impetus for the major reviews came from sources, particularly cash crises, other than either the National Curriculum or the OFSTED inspection system.

In the case of the United Synagogue Board of Religious Education, a major impetus came from the financial crisis in which the parent body the United Synagogue found itself in 1989, where it became clear that the cost of supporting the schools for which it is the foundation body was adding considerably to the financial crisis. The report however took on the issue of Jewish education as one not simply of financial exigency but as a central dilemma for the future of Jewish life in the UK. The foreword to the report was written by the newly installed Chief Rabbi, Dr Jonathan Sacks, who argued eloquently that at key moments when Jewish survival was at stake, it had always been initiatives related to education which had proved the turning point in Jewish survival (Note 11). This theme was to be amplified and promoted even more dramatically as the central issue for the very future of the present Jewish community in the UK.

In 1994, the Chief Rabbi published "Will Our Grandchildren be Jewish?" (Sacks, J (1994)), developing the arguments used in the foreword to the United Synagogue's report. (Note 12) It argued in particularly vivid terms that the current substantial continuing demographic decline of the Jewish community could be halted only by initiatives centered on Jewish education. This book was in turn the starting point for the launch of a very high profile and ambitious communal funding and development organization, Jewish Continuity. Jewish Continuity's initiatives began with full page advertisements in the "Jewish Chronicle" depicting the decline through intermarriage of the Jewish comm unity in an image of ranks of young Jewish people relentlessly marching over the edge of a precipice. It announced commitments to major initiatives to improve Jewish education, both formal and informal, and Jewish communal life (Note 13). A substantial component of these was a start-up establishment of a unit for research and quality development in Jewish education at a cost of over 31,000,000 Pounds Sterling. Further substantial funding for education development was to be made available through an open competitive scheme for grant awards to be allocated twice yearly to Jewish schools and educational bodies.

Alongside this, Jewish Continuity contributed substantially to the establishment of an ambitious new foundation body, designed to replace the United Synagogue Board, as recommended in the United Synagogue's report. The new body, the Agency for Jewish Education was set up with the goal of becoming a self-funding agency. Amongst the goals set out in its first strategic development plan was the development of an inspection service, including the preparation of Jewish schools for Section 13 inspection (Agency for Jewish Education (1994). Additionally a more long term target was the establishment of a new agreed syllabus for religious education.

The ZFET's review had identified additional problems relating to the system of having a series of two-year secondments from Israel for its Director of Education. A central theme for ZFET's report was issues related to the quality of Hebrew teaching and the lack of a common national curriculum frame work. Nevertheless, no mention was made of the existence of the National Curriculum framework for modern languages and the fact that it was legally compulsory for the secondary years. A conference held for ZFET Head teachers, heads of Hebrew and Jewish Studies and governors in May 1992 included a keynote speech on the implications of NC modern languages for the teaching of Hebrew. It was received with interest but no further initiatives were taken at that time either by ZFET or individual schools.

The emergence of the new Jewish Continuity funding structure together with personnel changes proved to be a decisive catalyst for refocusing the organization's energies on tackling the development of Hebrew teaching to take account of both National Curriculum and OFSTED criteria. A funding proposal was submitted to Jewish Continuity in April 1994 (Serra and Keiner (1994), proposing the development of a specific curriculum and assessment framework for Hebrew to be based on the model of National Curriculum modern languages, explicitly in order to enable schools to meet the challenge of having their achievements in Hebrew teaching assessed by OFSTED.

In the event, Jewish Continuity rejected the proposal as marking too radical a departure from traditions of Hebrew teaching, but ZFET proceeded with a modified version of the proposal by committing substantial funding from its own resources. With the prospect of OFSTED inspection imminent for its schools, the Head Teachers expressed enthusiastic support for the initiative. Pilot work in developing the curriculum approach was carried out in two schools, one of which underwent an OFSTED inspection in the Autumn of 1994. By the summer of 1995, following six months' drafting and consultation, the organization, now renamed the Scopus Jewish Educational Trust (Scopus) published curriculum frameworks for both Hebrew and Jewish Studies, both based very closely on the revised National Curriculum frameworks, including the specification of attainment targets, level descriptions and specific programs of study for each of the Key stages from 5-16 (Keiner, Korn, Serra and Frankel (1995a, 1995b). The consultation process revealed continuing strong support and commitment to adoption of the frameworks by the schools.

A third major Jewish communal body came to take an increasingly proactive role with Jewish schools in response to the emergence of the OFSTED system. This was the Education Committee of the Board of Deputies of British Jews (BD). The BD is a long-established representative body for the British Jewish community, its membership representing mainstream orthodox and reform synagogues and other communal bodies. Because it does include representation of non-orthodox religious groupings, it differs from both the major education foundations which are orthodox foundations, and it has therefore claimed and been given legitimacy in consultations with national bodies by reason of this wider degree of representation.

Over the years its education function has been primarily that of representing Judaism and

Jewish educational concerns to the non-Jewish educational world, for example developing training and curricular materials about Judaism for non-Jewish schools. It has also had an important role in negotiating with examination bodies and local authorities about providing for the observance of Jewish holy days for Jewish examination candidates and teachers. In practice, all its materials and pronouncements can be seen to contain no element which represents interpretations of Judaism and Jewish practice other than the orthodox.

With the advent of the National Curriculum, with its extensive programme of consultation at the stage of the development of the proposed curricula, three e BD came into increasing prominence on the national educational scene, as the DfEE's first port of call for consultation of the Jewish community. As the religious education initiatives referred to above came into prominence, the BD came to play a major role as the effective sole representative of Judaism on the national curriculum development body responsible for outlining model religious education syllabuses. The Director of Education of the BD was one of a new breed of Jewish community professionals, proactive and ready to play a high-profile role in promoting Jewish education and Jewish educational interests. Previously, Jewish community professionals involved in education have tended to be highly successful in promoting Jewish education through an unrivalled command of official procedures and informal consultative processes with central and local government education administrations.

### The emergence of a major initiative on the inspection of Jewish education

Once the OFSTED system of training inspectors had been established, the BD set out a initiative to influence and co-ordinate the selection of inspectors for the inspection of Jewish schools in general and of Section 13 inspections in particular. It began with the more traditional method of forming an invited working group drawn exclusively from educationists who were members of the orthodox community and within the United Synagogue's sphere of influence (Note 14). It also advertised in the major communal newspaper, the "Jewish Chronicle" asking any Jews who had qualified in OFSTED training to contact the BD in order to register as qualified inspectors with Jewish status. Three e BD as the result of its group meetings evolved an ambitious programme which ch could be seen as amounting to a major if inexplicit challenge to the two Jewish foundation bodies in seeking to become the most influential body in relation to quality control of Jewish schools.

It subsequently emerged that a more subtle process of religious vetting would be involved in the BD's proposal to establish a register of qualified inspectors with Jewish status. At a meeting of the Association of Governors of Orthodox Jewish Schools in April 1994, the Director of Education commented that "OFSTED inspection will be able to do for schools what heads and governors have wanted for years" (Note 15). He outlined the BD's intention to establish a training programme for inspectors of Jewish schools which he hoped would be the sole validated route recognised by OFSTED such schools. He envisaged that the religious credentials of inspectors to be involved in Jewish school inspections would be subject as part of this process to approval by the senior judge of the United Synagogue's ecclesiastical court.

The newly established OFSTED bureaucracy appeared to be as eager to embrace the BD's initiative as the Board itself was to establish it. Faced with the prospect of including inspections of up to a quarter of the existing Jewish voluntary aided schools in the first year of its operations, OFSTED's then Chief Executive established contacts with the office of the Chief Rabbi and the BD and was prepared to offer accelerated access to OFSTED training, for which there was a substantial waiting list to candidates approved by the BD.

In February 1995, the present Chief Executive of OFSTED gave the keynote ad dress at a conference of teachers called by the BD to promote awareness of the implications for Jewish schools of OFSTED inspection (Note 16). He stated that OFSTED looked forward to Jewish schools defining statements of religious values as a contribution to OFSTED's work on seeking to define what constitutes spiritual, moral, social and cultural values, suggesting that in mainstream schools there were insufficient initiatives of this kind. Much of the discussion at the Conference centered on the desirability of establishing an approved list of inspectors for Section 13 inspectors of Jewish schools. The Director of Education of the BD argued enthusiastically for inspections of Jewish schools not to be carried out by inspectors who were merely Jewish but by inspectors who were Jewish by practice and conviction, a view which was not universally endorsed by the meeting.

The BD subsequently obtained substantial funding, from Jewish Continuity, of over 310,000 Pounds Sterling to develop a framework for Section 13 inspections of Jewish schools, designed to parallel the published framework for OFSTED's Section 9 inspections. In doing so, it was emulating initiatives taken by the two major Christian voluntary school foundation bodies, the Church of England and Roman Catholic Diocesan authorities. The BD's initiative was as ambitious as that of the Scopus organization in formulating its curriculum proposals. In July 1995, BD issued the first draft of a very detailed framework (Note 17). Entitled "Pikuach" (Hebrew -- inspection), it adopted a novel approach to the interpretation of the legal responsibilities for Jewish religious education. The proposals were sent in confidential draft form to the Head Teachers of all Jewish schools, with a covering letter stating that Head Teachers were to have ownership of the proposals, although a wider process of consultation would be involved. The responsibility for religious education matters in voluntary aided schools in fact rests with the governors of each school, and to some degree with the foundation bodies which appoint them. The BD's stance was analogous to according ownership of quality control procedures for enterprises such as public utility companies to the chief executives of those companies.

The proposals assigned responsibility for reporting on whether the assemblies conformed to the legal requirements to the Section 13 inspection, although the law assigns them to Section 9. The proposals suggestions for the evaluation of pupils' spiritual and moral development went far beyond the scope of the equivalent criteria for the review of religious education in secular schools, as outlined in the OFSTED handbook. Additionally the proposals required Inspectors to take into account the "levels of Jewish commitment amongst the communal groups served by the school" and "any other relevant influences on pupils' behavior and Jewish values which are at play in the wider community and the school environment". These specifications constituted a significant departure from the generally firmly evidence-based approach of OFSTED criteria, because there is no readily available way in which such judgements could be made on other than a common sense speculative basis.

Additionally the requirement, made in the first drafts and subsequently removed, to consider the levels of Jewish commitment amongst all the school's Jewish teachers, not specifically those involved in Jewish religious education, brought to the proposals an approach to inspection not otherwise encountered in English educational practice. The final edition (BD (1996)) requires inspectors to take into account the degree to which teachers are in sympathy with the Jewish ethos of the school. Nevertheless, for the most part the BD proposal was very closely modelled on OFSTED's handbooks, and as such added up to by far the most searching and rigorous framework for quality control ever applied to Jewish education in England.

As these proposals came to fruition, they were challenged by new developments which had threatened the credibility and even the existence both of OFSTED and the new Jewish Continuity organization. There was a continuing and rising outcry from school staffs about the impact of OFSTED inspections, based on allegations that the documentation required by the inspections produced unacceptable overload. This came at a time when the Conservative administration, faced with an increasingly dismal public standing, was ready to make concessions to teacher unions which it had previously been determined to face down. The OFSTED process was subjected to a review, and a considerably slimmed down new Handbook produced, to apply to all inspections from April 1996. Subject specific inspection guidelines were replaced by generic curriculum criteria. However, new subject criteria (Note 18) were published and issued to OFSTED team inspectors, thus making the supposed slimming down appear perhaps more of presentation than substance. But the pre-existing separation between the Section 9 and Section 13 regulations was left untouched, even though consultations with

OFSTED inspectors had indicated their wish to have the anomalies clarified and at least some more decisive guidance on the boundaries between the two types of inspection.

Jewish Continuity itself became a subject of intense controversy inside and outside the Jewish community. A major television documentary made by the BBC as part of its prestigious prime-time "Everyman" series portrayed it as a almost sinister body bent on promoting Jewish separatism, inspired by three e advertising which had sought to sensationalize Jewish outmarriage. More sustained and damaging controversy bubbled up repeatedly within the Jewish community, focussing on the incompatibility of its claims to be a cross-community body, whilst quietly ensuring that all its major decisions and recipients were within the United Synagogue or other orthodox orbit.

It is not clear whether senior policy makers at OFSTED were aware of the fact that BD initiatives concerned with education were effectively becoming enmeshed within the "turf wars" amongst the various Jewish communal and professional organizations concerned with education. Senior OFSTED officials continued to appear at BD-organized events related to the development of "Pikuach", notably a consultative conference held to discuss its third draft, in November 1995 (Note 19), at which the President of BD referred to its claims to " work across cummunal boundaries and reach across the divisions" and to its "vibrant and proactive role in enhancing Jewish education". Thus from having previously been an organization largely confined to advocacy of Judaism and Jewish educational roles to the wider world, BD was now claiming a central, perhaps the central role in promoting Jewish education in the UK.

In March 1996, Jewish Continuity published a self-review (Note 20), based on substantial consultation across the Jewish professional and lay communities, which reflected the profound disquiets and conflicts raised by its ambiguous position, including its position in seeking to promote educational developments. It reported views that its interventions in education had been seen as aggressive, ignoring existing communal expertise, and that its decisions were thought by many to be taken privately by its Chairman and Chief Executive. The report proposed to remedy this by reconstituting the organization as a genuinely cross-communal initiative. It remained at the time of writing to be seen whether this could be achieved in a situation where Orthodox participants will accept only the legitimacy of their own authorities within any cross-communal initiative.

### OFSTED's first inspection findings on Jewish schools

The OFSTED system had by the start of the 1995-96 academic year been in full operation for two years, although the programme of primary inspections only began in 1994-95. Under the legislation, inspections of schools are required to take place once every four years. In practice, the full quota of a quarter of all primary schools which should have been completed has not been achieved for two reasons. Firstly, the number of inspectors so far successfully trained for primary schools and for special educational needs has not been sufficient to carry out the inspections. In addition, the independent free market system for awarding inspection contracts has resulted in OFSTED receiving no bids or only one bid for substantial numbers of schools.

By February 1996, three inspections had taken place of Jewish voluntary aided schools, two of secondary schools and one of a primary school. Of those schools, two of the secondary schools are grant maintained, one of them having a link to the United Synagogue, and the other two being independent Orthodox foundations. The primary school is part of the Scopus (formerly ZFET) network. All the Section 9 teams inspection included at least one Jewish inspector. In the case of the two London secondary schools, the Registered, or lead, inspector was Jewish, and there were additional team members who were Jewish. In the case of the primary school, there was more than one member of the inspection team who was Jewish. However, as the Director of Education of BD had pointed out, membership of Jewish ethnic credentials did not necessarily indicate knowledgeability about Jewish religious education and values.

Of all the schools, only the primary school had its Section 13 inspection take place at the same time as the Section 9 inspection. The governors appointed a single inspector who is an OFSTED-trained deputy head teacher, with specialist training in Jewish religious studies, whose school is a member of the same foundation body as the inspected school. In the case of one secondary school, the inspection took place separately from the Section 9 inspection, and was conducted by two inspectors, both members of the orthodox Jewish community, one of whom is an OFSTED accredited inspector who also serves as a local authority inspector, and one of whom is an OFSTED accredited lay inspector.

In the case of another secondary school, the Section 13 inspection took place eight months after the completion and publication of the Section 9 inspection, and in the next academic year. This inspection was thus in breach of the DfEE regulations which state that the Section 13 inspection must be conducted in the same academic year. The general DfEE regulations also state that in the case of a Section 9 inspection, inspectors must not have had any significant prior connection with the school in either a personal or a professional capacity. In the case of this school, the school's governors awarded its Section 13 contract to a gentile inspector who was formerly the religious education adviser for the local authority of which the school was a part before the school obtained grant maintained status. This would appear to raise further issues about the procedures governing the two types of inspection, since it would not appear that there have been any consequences a rising from the apparent breaches of the regulations.

The inspection teams of the schools which have completed a Section 13 inspection have thus been different both in terms of composition and mode of inspection. No Section 13 inspection to date has used a set of published criteria to work to which was specific to Jewish education. Indeed, in no case has any set of criteria used been explicitly identified. In no case was the Section 13 inspector solely responsible for reporting on the spiritual, moral, social and cultural dimension of the school, or for the school's achievements in Hebrew teaching. In fact, in the case of all the Jewish schools inspected so far, there are paragraphs on pupils' personal development and behavior in the Section 9 report covering the social, moral, spiritual and cultural dimension, based on the criteria specified in the 1993 OFSTED handbook. The equivalent Section 13 reports, with one exception, have paragraphs which are largely confined to statements about the extent to which spiritual, moral, cultural and social issues are encountered in the school's assemblies and religious studies programmes. Thus these inspections already demonstrate that, in practice, judgments about the school in general and about its Jewish ethos in particular appear to be being made in a different way from what was intended by the legislation.

The Board of Deputies' initiative "Pikuach" (BD 1995(a), 1995(b), (1996)), referred to above, is making enthusiastic claims to meet the need for clear criteria. It certainly offers a comprehensive descriptive framework, but its criteria for evaluation could be said to beg the question, since it leaves it to each school to specify which criteria are to be used for the purposes of inspecting the content of Jewish Studies courses. Thus, the situation, referred to above, in which one school does not offer preparation for any external Advanced Level syllabus examination cannot be judged a failure or a serious weakness, because the school itself makes a judgement that the existing examinations do not match its self-chosen criteria for teaching Jewish Studies. A basic principle of OFSTED is to make judgements against criteria which are either explicitly stated within laws and regulations, or within the legally compulsory NC subject documentation. Thus the claim of Pikuach to legitimacy for inspection purposes within an OFSTED framework appears to be difficult to reconcile with that principle.

### Inspection findings on the ethos of Jewish schools

In the case of all the schools, we need to look to the Section 9 inspection report for judgements about the extent to which the schools are achieving a Jewish ethos overall. In the case of both the secondary schools, the Section 9 inspectors commented on the relative lack of integration between the secular studies of the school and its Jewish life. In the case of one secondary school, the stark comment was that

> ...most teaching misses valuable opportunities to contribute to pupils' spiritual development. Likewise, outside Jewish studies and modern Hebrew, there are few references to Jewish culture in the curriculum, with the result that Jewish matters are separated from secular matters. The school should consider whether this situation accords with its ethos. (OFSTED (1995a) para 33)

The Section 9 inspection of the second secondary school reported that

> The curriculum makes a variable contribution to pupils' cultural development. In most subjects the content is restricted to white western cultures. Modern Hebrew plays a role in reflecting and affirming Jewish identity, values and experiences; some Holocaust literature is read and discussed in English; Jewish musical styles are studied and performed, alongside culturally and stylistically varied musical traditions; and in art there are incidental references to Jewish craft and design traditions and their contribution to culture in a variety of contexts. However, the potential for Jewish exemplars in all areas of the curriculum is not fully realized. Pupils generally do not appreciate deeply enough how other societies function and pupils awareness and appreciation of cultural diversity is limited.(OFSTED (1994c, para 39))

The primary school's Section 9 report, while praising the positive impact of the school's Jewish life on the school as a community, made similar points about the relative insulation of Jewish ethical perspectives from those of the curriculum as a whole:

> ...prayer is an important feature of each day, restating and celebrating three e school's values and beliefs. There is scope across the curriculum to address spiritual and moral issues more directly and to promote greater levels of curiosity and a sense of discovery amongst the pupils. Attitudes to work and to the life of the school are positive. There is a strong Zionist flavour throughout the school and the children are taught Hebrew as a second language. However, the pupils need to explore more fully the variety of cultural traditions both within their own and the wider world. (OFSTED 1994b)

## Dilemmas of inspecting Hebrew teaching

Further common findings of the inspection reports related to the teaching of Hebrew, reported on as a modern foreign language as part of the Section 9 report. Although Hebrew reading is a major component of Jewish studies, neither the Section 9 nor the Section 13 inspection reports of the secondary schools addressed the issue of the effectiveness of the modern Hebrew teaching in contributing to preparing pupils for those needs. In the case of both the secondary schools, the Section 9 reports commented that the Hebrew department needed a closer relationship with the separate modern languages department. In both schools, comments on the status and quality of Hebrew teaching reflected a mixed verdict.

Although achievements in public examinations were above expected national standards, and pupils benefited from teachers who were native speakers, there was evidence of underachievement by lower ability pupils, and of a lower status being accorded to Hebrew as an option beyond the first two years of the school. In spite of its importance in relation to the schools' ethos as Jewish schools, the schools offered Hebrew as an examination subject only beyond the first two years of the secondary phase. In the case of one of the schools, it was criticised for offering Hebrew, for pupils who wished to take both French and Hebrew, only as a course to be taken outside school hours.

The reports on both secondary schools reflected variations in the quality of teaching and learning, with a significant minority of lessons showing evidence of poor organization. In one school, no pupil below sixth form level was observed to speak Hebrew spontaneously. In neither secondary school was the use of information technology incorporated into Hebrew teaching as required by NC, and pupils did not make sufficient use of dictionaries and glossaries. Both secondary reports commented on insufficient provision to meet the needs of pupils with learning difficulties.

In the case of the primary school, the Section 9 report commented favorably on the Hebrew teaching offered, and the Section 13 report specifically considered the extent to which it enabled the pupils to tackle religious texts. The latter report identified lack of liaison between the Hebrew and Jewish studies departments as contributing to mismatch between pupil capability and teacher expectations.

**Inspection findings on the quality of Jewish Studies**

In terms of the specific quality of religious education in Jewish schools, there are now three Section 13 reports published (OFSTED, 1994b; 1994d; 1996) although as shown above, the Section 9 reports did address the impact of aspects of religious education across the whole of the curriculum offered by the school. All the reports commented substantially favorably on the Jewish studies curricula of the schools. All commented on the positive effect of the programmes of Jewish teaching offered on the pupils' social and moral development.. On the case of one secondary school and the primary school they also reported on the pupils' knowledge of Jewish prayers and practices, identifying substantial knowledge of texts.

The Section 13 report on the second secondary school contained many highly complimentary findings, but also more surprising ones, such as the fact that it does not conform with legal requirements for collective worship, that its pupils do very little written work in Jewish studies, that its GCSE results in Jewish Studies are substantially lower than in the great majority of secular subjects, with those of girls showing a very substantial decline in the last year. It reported that by choice the school does not offer any Advanced (University Entrance) Level examination courses in Jewish Studies. There appeared to be no attempt in this report to evaluate the pupils' knowledge of Jewish texts or prayers and other rituals. Among its most complimentary findings were those on the success of its Informal Education program of Jewish studies, which includes organized periods of study in Israel, study weekends and other activities in and out of school. Nevertheless the report indicated that only a small minority of the school's 1400 pupils participated in the programme. The report commented that the school had no objective system designed to measure the success of its objectives of increasing commitment to Judaism, Israel and Jewish life.

In fact, in all cases, the Section 13 reports drew attention to the relative lack of in-house monitoring and evaluation of the quality of Jewish education. All the reports comment on the lack of effective whole school assessment policy in Jewish studies, with considerable variations of assessment and marking practice. The primary school report indicated that no records were being kept of progress in Jewish studies.

The messages in the reports so far do much to confirm and extend the analyses presented in the earlier reports of the United Synagogue and the Scopus foundation bodies. Those reports primarily focused on the need to build better structures and mechanisms for those bodies, and on the need for a major program of general in-service training. However, it would seem that the enthusiasm which the Heads of the Jewish schools are showing for the establishment of published curriculum, assessment and inspection systems specific to Jewish Studies and Hebrew, owes much to the advent of the OFSTED inspection era with its system of published criteria, quality control procedures and published reports.

**Reporting inspection findings in the Jewish community press**

It is also an additional measure of the impact of the new inspection system that it provides a new focus for discussion of the performance of Jewish schools in the Jewish press. In recent years, the "Jewish Chronicle" has regularly published features summarizing the GCSE and A Level achievements of the various Jewish schools (Note 21). However, although the results of NC assessments at Key Stage 1 and Key Stage 3 have been published for several years, they have never been reported on in the Jewish or local press. The publication of OFSTED reports has attracted coverage, and the report in the Jewish Chronicle on one secondary school's OFSTED report highlighted criticisms made of the teaching of modern Hebrew (Note 22). The only mention of the primary schools in the OFSTED report referred to the inspectors' commendation of Hebrew and Jewish studies teaching. There is evidence of growing attention to achievements in these subjects, with the appearance of an editorial in the "Jewish Chronicle" in the same week as its reporting of Jewish schools' secular examination successes referring to the failure of the schools to reach the levels of achievement in Hebrew and Jewish studies required by the community (Note 23). Nevertheless, the fact that schools are able to set their own timetable for Section 13 inspections can mean that the attention of the press is avoided. The report on the school which had its Section 13 report published in the following academic year to is Section 9 report received no mention in the Jewish press, although it contained what might be thought to be some newsworthy revelations, as referred to above. This lack of press coverage was presumably due to the fact that inspection reports on the school were considered old news.

### Responses to inspection by governors and foundation organizations

The legislation on OFSTED inspections defined how schools must respond to both Section 9 and Section 13 inspection reports. It required the governors of each school to submit to OFSTED and publish to parents a separate action plan for each report, detailing their intended response to the key issues for action identified by the inspectors, within forty days of its publication. As already indicated above, for Section 9 reports of Jewish schools, this has in practice covered substantial aspects of the school's distinctively denominational practice, noticeably the teaching of Hebrew. In practice, unified action plans by Jewish schools have included responses to both the Section 9 and Section 13 reports (Note 24). It is not widely appreciated that governors of schools in England and Wales, community volunteers who have official responsibility for the curriculum and policy management of schools, have in the past had little access to direct evaluative evidence about the achievements of their schools, other than the results of external examinations and. latterly, the results of the externally monitored and marked tests which NC requires at ages 7, 11 and 14 for secular subjects. There has not previously been any consistent and reliable source of evidence about the efficacy of a particular school's Jewish Studies or Hebrew programme, and most governors of Jewish schools will readily acknowledge that they know little or nothing of what is achieved beyond what they can deduce from parental comments or public presentations by the school. The advent of OFSTED reporting adds dramatically to the base of evidence which is available to them.

Governors, head teachers and staff are now having to debate and agree responses to inspection reports, which may include responses related to school policy and practice on curriculum, resources and assessment. Those responses must ultimately derive from the teaching staff concerned, and it is now clear even with only a small number of inspection reports so far published th at the impact on them in terms of expectations and accountability will be considerable. Many responses will need to be at the level of the whole school, where such matters as resource allocation and assessment policy may need to be reviewed. A further major impact must therefore be in increasing the integration of Jewish studies and Hebrew teaching into the centre of school development as a whole.

### Parental interest in inspection reports

Whether this new level of accountability will have any lasting impact on parents remains to be seen. The very fact that the Section 9 and Section 13 reports are published separately may tend to lessen parental focus on the inspection verdicts on the specifically Jewish dimension of the school's achievements. While parents receive free of charge summaries of both reports, an indication of levels of parental interest can be derived from the number of parents and others being prepared to pay for full copies of reports, for which schools are allowed to charge. The demand for full reports for Section 9 inspections has been substantially higher than for full Section 13 reports. In only one school, copies of the complete Section 13 report have been provided to all parents, when the Section 9 report has been distributed to them as a summary, as required by the regulations. This suggests some particular motivation on the part of the school, perhaps connected with building parental support for desired policy initiatives, since the expense of duplicating the report must have been a significant budgetary decision taken by the governors and senior management staff.

Parental reasons for choosing a Jewish school are complex, including their assumptions about whether their children are likely to do better in secular subjects at Jewish schools, as well as considerations of their desire to foster their children's commitment to Judaism, and their perceptions of the peer groups their children might meet in non-Jewish schools . It is clear that the popularity of Jewish schools owes much to their high achievements in secular studies. Recent demographic research on the Jewish community suggests that only a small minority of the community actually practises orthodox Judaism (Note 25). The reports as circulated have included in the cases of some schools some very substantial criticisms in relation to both secular and Jewish studies. There is as yet little evidence that reporting on the quality of Jewish education and Hebrew will affect parental decisions for the vast majority of parents. However, it will certainly heighten awareness of what their children are and are not achieving in this field.

**Notes**

Elements of an earlier version of this material were previously presented at the Conference of the International Sociological Association Sociology of Education Research Committee, "Educational Knowledge and School Curricula: Comparative Sociological Perspectives", The Hebrew University, Jerusalem, December 27th 1995.

1. Minutes of a meeting on the Inspection of Jewish Schools, Board of Deputies of British Jews Education Department, 6th February 1994. Her Majesty's Inspector Mr R Long reported that there are an additional forty seven known Jewish independent schools, which HM Inspectorate service. Further applications for state-aided status are currently in the pipeline for at least five further Jewish schools, three of which are from reform or liberal Jewish bodies, and two from orthodox bodies. All but one are for the outer London suburban areas.

2. The oldest Jewish school in England, the Jews' Free School (JFS) comprehensive, formerly the Jews' Free School, dates back to 1817 (Gartner (1960) p.221).

3. ibid., p.22. There was opposition to the payment of grants to religious schools in general by some Liberal nonconformists at the time of the establishment of the state aid system established in 1870, with additional objection to support for non-Christian religious education. There was also opposition by nonconformists to religious education in secular schools, and it was open to the School Boards established by the 1870 Education Act to decide whether or not it was to be included.

4. Hyman & Ohrenstein (1993) cited four nursery schools, six primary schools and four secondary schools as being under the aegis of the ZFET. Two of the nursery schools and one of the primary schools are independent non-state aided schools.

5. Although by far the most influential organization in Jewish education, three e United Synagogue is directly responsible for only five of the twenty four state-aided Jewish schools.

6. It currently runs withdrawal classes in Jewish religious education at two major prestigious independent schools in London which have very substantial numbers of Jewish pupils.

7. In 1995, the Hasmonean High School, a Jewish comprehensive school, achieved the highest percentage for all comprehensive schools in England and Wales of A and B grades in the GCE Advanced Level examinations, and the sixth highest percentage of all state schools, including selective schools. The JFS comprehensive school achieved forty fifth place in the percentage rankings for state schools for A and B A Level grades, and the King David High School Liverpool achieved 178th place nationally. Rankings in the previous year were: Hasmonean High, fifteenth, JFS, twenty-second and King David High, Liverpool, fifty-third.

8. Jewish Educational Development Trust (1992), known as the "Worms Report", after its Chairman, Mr Fred Worms, was the United Synagogue's review; Hyman & Ohrenstein, op. cit., was the ZFET's review.

9. The core subjects are: English, mathematics and science. The foundation subjects are: technology, history, geography, art, music, physical education and, for pupils over 11, a modern foreign language.

10. The somewhat complex arrangements for Section 9 and Section 13 reporting on the spiritual, moral, social and cultural aspects of the school are set out in DfEE Circulars 7/93, Appendix B and 1/94, Para. 134. There is some ambiguity between the positions set out in the two documents, with Circular 7/93 stating in Appendix 6 Paragraph 6 that "inspection for a school which offers denominational education cannot cover this aspect, although it must cover the moral, spiritual, social and cultural development of pupils across the whole range of the school's activities". On the other hand Circular 1/94 Para. 134 states, "The Registered Inspector has the duty...to report on the spiritual, moral, social and cultural development of pupils in all schools, but in [denominational schools] that duty is limited to noting that the school meets the requirements of the law to provide RE and a daily act of collective worship. The Registered Inspector is not concerned with the content of such provision."

11. JDT (1992) op. cit. pages i-ii.

12. Sacks (1993) particularly at pages 34-48 and 104-111.

13. Jewish Continuity (1994). An initial outline of Jewish Continuity's goals and strategy was previously given in Sacks (1994) pages 106-111 and 117-123

14. Minutes of a Meeting of the Association of Governors of Orthodox Jewish Schools, 23rd January 1994. Notes of presentation by Mrs Syma Weinberg of Jewish Continuity.

15. Presentation by Mr Laurie Rosenberg, Director of Education of the Board of Deputies of Jewish Schools, 24th April 1994, Meeting of the Association of Governors of Orthodox Jewish Schools.

16. Meeting of Jewish Teachers' Forum on "OFSTED and the Jewish School", organized by the Education Department of the Board of Deputies, 1st February 1995

17. Board of Deputies of British Jews Education Department (1995a)

18. See for example OFSTED (1996a)

19. "Pikuach" Board of Deputies Education Department Consultation Conference, 20th November 1995

20. Reported in the Jewish Chronicle, 15th March 1996, pages 1 and 25.

21. Cf. Jewish Chronicle 25th August 1995

22. Jewish Chronicle 3rd February 1995

23. Jewish Chronicle 25th August 1995, Second leader.

24. For example, Simon Marks Jewish Primary School (1995)

25. See JEDT (1992), Section 1, para 1.1

## References

Agency for Jewish Education (1994). Target 2000.5 year Plan 1995-2000 (mimeo)

Alderman, G (1989) London Jewry and London Politics 1889-1986. London: Routledge.

Board of Deputies of Pikuach. Inspecting Jewish Schools. The Framework. First Draft (mimeo)

Board of Deputies of British Jews (1995a) Pikuach. Inspecting Jewish Schools. The Framework. First Draft (mimeo)

Board of Deputies of British Jews (1995b) Pikuach. Inspecting Jewish Schools. The Framework. Third Draft Version 2.1, November 1995.

Board of Deputies of British Jews (1996) Pikuach. Inspecting Jewish Schools. The Framework. 22nd February 1996.

Gartner, L (1960) The Jewish Immigrant in England 1870-1914 London: Simon Publications.

Hyman, B & Ohrenstein, A (1993). Blueprint for Progress. An examination of the present structure of the ZFET with recommendations for the future (mimeo).

Jewish Continuity Strategic Direction.(1994) 5 Year Goals and 1995 Programme. 22nd December 1994 (mimeo)

Jewish Education Development Trust (JEDT) (1992). Securing Our Future: An Inquiry into Jewish Education in the Development Trust United Kingdom.London: Jewish Education Development Trust.

Keiner, J, Korn, E., Serra, D. & Frankel, J. (1995a) Scopus National Jewish Studies Curriculum. London: Scopus Jewish Educational Trust

Keiner, J.,Korn, E.,Serra, D. & Frankel, J. (1995b) Scopus National Hebrew Curriculum. London: Scopus Jewish Educational Trust

Sacks, J. (1994). Will We Have Jewish Grandchildren? Jewish Continuity and How to Achieve It. London: Valentine Mitchell.

Serra, D. & Keiner,J.(1994). New Life for Hebrew Teaching. Proposal for funding to Jewish Continuity (mimeo)

Simon Marks Jewish Primary School (1995). OFSTED Action Plan 1994-96 (mimeo)

**Official Publications**

Great Britain (1988) Education Reform Act Chapter 40. Part I, Chapter 1, 1(2)

Great Britain Circular 7/93. Inspecting Schools: A Guide to the Inspection Department for

Provisions of the Education (Schools) Act 1992 in England. Education and Science(DfE) (1993)

Great Britain Circular 1/94. Religious Education and Collective Worship Department for . Education(DfE) (1994)

Office for Standards Handbook for the Inspection of Schools in Education London, HMSO (OFSTED)(1993)

Guidance on the Inspection of Secondary Schools. The OFSTED Handbook, London, HMSO (1995b)

Guidance on the Inspection of Nursery and Primary Schools. The OFSTED Handbook. London, HMSO (1995c)

Guidance on the Inspection of Special Schools. The OFSTED Handbook. London, HMSO (1995d)

Inspection Resource Pack. The OFSTED Handbook. London, HMSO (1995e)

Primary Subject Guidance; London, Office of Her Majesty's Chief Inspector of Schools in England (1996a)

**OFSTED Inspection Reports**

OFSTED (1994a) Report of Inspection under Section 9 of the Education (Schools) Act. Simon Marks Jewish Primary School. Inspected 3rd-6th October 1994

OFSTED (1994b) Report of Inspection under Section 13 of the Education (Schools) Act 1992. Simon Marks Jewish Primary School Inspected 3rd-6th October 1994

(OFSTED 1994c) Report of Inspection under Section 9 of the Education (Schools) Act. Hasmonean High School, Inspected November 21st-25th 1994

(OFSTED 1994d) Report of Inspection under Section 13 of the Education (Schools) Act 1992. Denominational Religious Education and Collective Worship.Hasmonean High School. Inspected 4th-8th December 1994.

OFSTED (1995a) Report of Inspection under Section 9 of the Education (Schools) Act. Jews' Free School, Inspected March 9th-15th 1995

OFSTED (1996b) Report of Inspection under Section 13 of the Education (Schools) Act. JFS. Inspected 20th-22nd November 1995.

## About the Author

Judy Keiner is Senior Lecturer in Education, Department of Education Studies and Management, University of Reading, England. She is also a school inspector. Her undergraduate degree is in English literature, and has postagrduate degrees in sociology and information technology. She taught in primary and secondary schools and in further and higher education and has worked for over 25 years in teacher education. Her current research interests focus on Jewish education and on the use of the World Wide Web by children. Samples of her current work may be reached via her Web page at http://www.reading.ac.uk/~veskeinr/.

Email: J.C.Keiner@reading.ac.uk

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, Az 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 6** **April 3, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS
ARCHIVES.Permission is hereby granted to copy any
article provided that EDU POLICY ANALYSIS
ARCHIVES is credited and copies are not sold.

---

# The 1976 Illini:
# Sweet Memories of Alma Mater

**Diya Dutt**
**University of Illinois--Urbana, Champaign**

*dutt@uiuc.edu*

**Abstract:**

The purpose of this article is to explore the attitudes of graduates of the class of 1976 from the University of Illinois toward their alma mater over a period of fifteen years. The central question addressed in this article is: How do former students feel about their educational institution as time passes? Early research suggests that students' attachment to their educational institution becomes weaker with the passage of time. This panel data on alumni attitudes towards the academic environment indicates that contrary to evidence from past research, students developed a stronger attachment towards the educational institution with passage of time. A similar positive pattern was evident when examining the attitude towards the program major. It is possible that better experiences in the real world have made the alumni comprehend the quality of education they received at the University of Illinois. Also, favorable disposition toward one's institution seems to be, to a very considerable extent, the college's contribution to the intellectual development of the student.

The purpose of this article is to explore the attitudes of the graduates of the class of 1976 from the University of Illinois toward their alma mater over a period of fifteen years. The central question addressed in this article is: How do former students feel about their educational institution as time passes? Assessing how well students regard both the university and the education they receive is important for evaluation and planning purposes. This article explores graduates' satisfaction with their educational experience and assesses how positively respondents feel toward the university, their major, and the preparation provided by their majors for their careers. Early research suggests that students' attachment to their educational institution becomes weaker with the passage of time. Does the students' attitude toward the institution change differentially once they graduate from the University?

Few longitudinal studies spanning a decade or more of the formation of opinion by graduates toward academic institution have been undertaken in higher education research. The data for this paper originated from a panel study of the class of 1976 graduates from the

University of Illinois who were interviewed at four points in time. Panel studies like this cost a great deal of time and money, but they help in building a rare data base for educational institutions which permits an analysis of student trends for usage in program review and planning.

## Literature Review

Alumni research is crucial for assessing the long range benefits or detriments of college academic experience. The hallmark of a good University is the product -- the alumni (Spaeth, 1981) and they are an important part of higher education's constituency (Pace, 1979). However, literature in the field of alumni research has been meager until today. A delay in alumni research can adversely influence educational management issues like program review, curriculum planning, student assessment, resource allocation, and career counseling (Melchiori, 1988; Moden & Williford, 1988). Following alumni through their lives and focusing on demographic characteristics, attitudinal issues, and career patterns can help unravel the motivational forces of alumni as providers for their institutions (Melchiori, 1988; Stover, 1930).

Alumni research gained momentum after the 1930s because the economic depression stimulated systematic objective inquiries into the plight of college graduates (Pace, 1979). Two studies were conducted by the University of Minnesota and the U.S. Office of Education during the years of the Great Depression to determine the economic status of college alumni. The Minnesota study found that job opportunities for college graduates were markedly limited during the Depression years. However, more than sixty percent of the students got jobs in the same field as their college specialization. The average yearly salaries were low for men and uniformly lower for women (Pace, 1979). The results of the Minnesota survey were confirmed by a nationwide study of college graduates reported by the U.S. Office of Education (Pace, 1979). The study encompassed college graduates from 31 different colleges and universities during the years from 1928 to 1935, and confirmed the hardships faced by college graduates during the Depression era (Pace, 1979).

Following the Second World War, a landmark study of college graduates was conducted by the research division of Time Magazine (Pace, 1979). The Time study was a national sample of all college graduates whose names were obtained from 1200 degree-granting colleges and universities in the late 1940s. The survey included questions about the economic and occupational status of the alumni, their attitudes toward college and their involvement in civic, cultural and political affairs. The study revealed that a majority of the students attached a high value to their college and asserted that they would go back to the same institution from where they received their degrees.

Following the Time survey, the next alumni study of national scope was done in 1963 at the Survey Research Center of the University of California, Berkeley (Pace, 1979). The scope of the study went beyond job opportunities for students after graduation, delving into attitudes about their own education, its benefits, and also their involvement in a variety of civic and cultural activities. The major importance of this study was that it concentrated on the lives of men who had graduated with a major in one of the traditional liberal arts fields, i.e., the social sciences, humanities, literature, and the arts (Pace, 1979).

Another survey of nationwide scope was conducted by the National Opinion Research Center (NORC) in 1969. This included samples of alumni from the graduating class of 1961 from 135 colleges and universities. The result of the study was reported in a book written for the Carnegie Commission on Higher Education (Spaeth, 1970). The authors wanted to know how members of the class of 1961, after graduating a decade ago, assessed the performance of their alma mater. Among other issues, they wanted to ascertain the attitudes of former students toward their University. In their study, they found that nostalgia for their alma mater was not overwhelming among the alumni (Spaeth, 1970). Those who had a strong attachment to their college had declined in number a decade after they graduated from the University. It could be that experience in the outside world or the mere passing of time had moderated strong positive feelings toward the university (Spaeth, 1970).

Another study investigated the effects of various aspects of the academic environment on students' satisfaction with the college experience (Rich & Jolicouer, 1978). Data for this study was collected from 12 colleges and universities in California in the fall and winter of 1975-76 (Rich & Jolicouer, 1978). The authors found that longer tenure in college is negatively associated with positive rating for institutions. Students become disenchanted during the course of their stay in college, and high expectations they had from high school give way to realities of hard work, less success and difficulties with peers and faculty (Rich & Jolicouer, 1978). Interestingly, they also observed that students at public colleges rate their school less highly than those at private institutions (Rich & Jolicouer, 1978).

### Research Hypotheses

This article explores student attitudes toward the University of Illinois and major Programs of Study over a period of fifteen years. Based upon the literature pertaining to alumni attitudes and higher education, the research hypotheses developed for this paper are:

- Strong positive feeling toward the college declines substantially with the passage of time.
- Attitude towards program major becomes more positive with better experience in the job market.
- Positive disposition towards the educational institution is a function of the University's contribution to the intellectual development and of the perception of faculty concern for student needs.

### Research Design

The University of Illinois has conducted surveys of its graduates since 1973. The class of 1976 is unique because it has been surveyed four times at intervals of one, five, ten and fifteen years. The survey included measures to assess students' post-graduation employment history, further educational achievements, attitude toward the University and major Program of Study, and satisfaction with the quality of instruction and course offerings. The University Alumni Association maintains a database containing demographic information of all University alumni. This file provides information for each alumnus including home address, major curriculum code, degree awarded, sex, ethnic code, campus location, graduation month, birth date, and social security number.

This article is based on data collected in four waves (1977, 1981, 1986, 1991) through a 29 item, self- administered mail questionnaire. This was a population survey of graduates of the class of 1976 from both Urbana and Chicago campuses (N=12,854). A packet of materials, including a cover letter signed by the President of the University, the instrument, and a pre-addressed stamped envelope was mailed, using first class postage, to each respondent. Two follow up mailings of non-respondents were done at an interval of three weeks to enhance the response rate. This study is based on the pool of graduates who have participated in all four surveys (N = 2,306) (Note 1).

### Statistical Design

Repeated Measures Analysis was used to analyze alumni's emotional attachment to the University and attitude toward major Program of Study over time. (Please refer to the Appendix for detailed observation on the choice of statistical design). Cronbach's alpha was utilized to construct two indexes to measure program satisfaction and faculty guidance. The coefficient Alpha is based on the inter-item correlation, which helps decide whether a group of items should be added together to form a scale or index. Ordinary Least Squares (OLS) regression procedure was used to assess the impact of program satisfaction and faculty guidance index on the attitude towards the University. The Stepwise model selection procedure was used, where at each stage a test was made of the least useful predictor.

### Discussion of Findings

#### Sample Characteristics

The sample consists of 1469 males and 837 females. The mean age of the male respondents at the time of graduation was 25.43 years, versus women, which was 25.85 years. In the panel, 62.6 percent of the students were baccalaureates, 24.5 percent received a Masters degree, 6.4 percent received doctoral degrees, and another 6.5 percent received a professional degree from the University. Characteristics of sample respondents by age, gender, campus location, geographical site, and degree level are provided (Table 1). As far as age distribution and geographical location was concerned, there was no difference between the panel respondents from the original pool. However, more men responded in all four surveys compared to women, and the sample also had more students from the Urbana-Champaign campus than the Chicago branch. In terms of degree level, there was a higher percentage of respondents with doctoral degree in the sample, and only a few professional degree holders returned surveys compared to the original pool.

Table 1

CHARACTERISTICS OF SAMPLE RESPONDENTS BY AGE, GENDER, CAMPUS, GEOGRAPHICAL LOCATION, AND DEGREE LEVEL

| Variables | Original Sample (N=12,854) | Returned Sample (N=2306) |
|---|---|---|
| Age of Respondents (Mean Years) | 25.6 | 25.6 |
| Gender (in percent) | | |
| Male | 59.9 | 63.7 |
| Female | 40.1 | 36.3 |
| Campus (in percent) | | |
| Urbana | 69.3 | 82.6 |
| Chicago | 30.7 | 17.4 |
| Location (in percent) | | |
| Illinois | 83.0 | 80.5 |
| Outside Illinois | 17.0 | 19.5 |
| Degree Level (in percent) | | |
| Bachelors | 62.4 | 62.6 |
| Masters | 24.4 | 24.5 |
| Doctoral | 5.9 | 6.4 |
| Professional | 7.3 | 6.5 |

## Alumni Attitudes Toward The University

What was the reaction of the 1976 alumni toward the University in which they received their degree? In this section of the article, we used four dependent variables, the attitude towards the University (Note 2) surveyed at four different points in time in a repeated measures analysis. Table 2 compares the reactions of the alumni over a period of fifteen years. The multivariate test (Hotelling- Trace=0.055) was significant at the .0001 level (F=43.52, degree of freedom =3, p = .0001) which meant that there was substantial change in the level of attachment towards the alma-mater over time. In other words, strong positive feelings by the alumni toward the college kept rising over a period. The Univariate test also shows significance at the .0001 level (F=49.69, degree of freedom= 3, p = .0001).

The overall statistical difference found among the attitudinal measures leads us to determine which specific time condition was responsible for contributing to this significance. In this repeated measures design, where a single group of subjects was measured at four points in time, we did a set of repeated contrasts. This was done to investigate whether there were significant differences at adjacent points in time. An analysis of variance was performed on the contrast variables, which represent the difference of mean between the attitudinal variable measured in 1977 with subsequent time periods. The results presented in the last column of Table 2 show that there was a substantial strengthening of positive feeling from former students toward the University over a period of fifteen years. The intensity reached its peak ten years after graduation but leveled off slightly after fifteen years.

Table 2
REPEATED MEASURES ANALYSIS OF ATTITUDE TOWARDS THE UNIVERSITY FOR THE CLASS OF 1976 OVER FIFTEEN YEARS

| Dependent Variables | Mean | Standard Deviation | Test of Contrast (1) |
|---|---|---|---|
| Attitude Towards University | | | |
| 1977 | 3.503<br>(N=2290) | 0.604 | |
| 1981 | 3.616<br>(N=2290) | 0.560 | F=82.38, df=1, p=.0001* |
| 1986 | 3.647<br>(N=2295) | 0.528 | F=117.10, df=1, p=.0001* |
| 1991 | 3.601<br>(N=2298) | 0.558 | F=47.40, df=1, p=.0001* |

| Multivariate Test | Univariate Test |
|---|---|
| Hotelling Trace=0.055, F=43.521, df=3, p=.0001* (N=2257) | F=49.69, df=3, p=.0001*<br>Greenhouse-Geisser I=.9299 (2)<br>(N=2257) |

1 The last column indicates the contrasts which represent the difference of means in 1977 with subsequent time periods.
2 The assumption of sphericity is tenable.

* Significant at .001 level.

## Positive Feelings Toward Program Major

In Table 3 we discover how the alumni rate their major Program of Study over a period of time. Positive strong feelings toward the major field of study were ascendant over a period of fifteen years. Repeated measures analysis was again used to gauge the intensity of feelings of alumni toward their major. The multivariate test (Hotelling- Trace=0.00929) was significant at .0001 level (F=6.955, degree of freedom= 3, p= .0001) which meant that there was an overall significant positive effect over time toward the major field of study by the alumni. The Univariate test also showed significance at the .0001 level (F=7.97, degree of freedom= 3, p = .0001). Again, since an overall difference was found, we wanted to determine which specific time period differed in the analysis. The analysis of variance for the contrast variable presented in last column of Table 3 revealed that there was a significant difference in feeling towards the major program of study over a period of ten and fifteen years. However, there was no appreciable change in response between 1977 and 1981 towards the major field of study (Table 3). It could be that a better experience in the post graduate world would have made the alumni realize the excellent quality of education received at the University of Illinois, which in turn strengthens positive reactions to major field of study over a period of time.

This finding is contrary to what past research indicates in general about alumni behavior (Rich & Jolicoeur, 1978; Spaeth, 1970). These studies on student attitudes toward academic environment indicate that in general, even though students are satisfied with their college, there is an erosion of strong positive feelings over time toward the university. It is interesting to note that one group of scholars (Rich & Jolicoeur, 1978) has indicated that students at public colleges rate their schools less highly than those at private institutions. In this respect, our finding is significant because the University of Illinois is a major public University.

Table 3

REPEATED MEASURES ANALYSIS OF ATTITUDE TOWARD MAJOR PROGRAM OF STUDY FOR THE CLASS OF 1976 OVER FIFTEEN YEARS

| Dependent Variables | Mean | Standard Deviation | Test of Contrast (1) |
|---|---|---|---|
| Attitude Towards Program Major | | | |
| 1977 | 3.345<br>(N=2284) | 0.704 | |
| 1981 | 3.360<br>(N=2294) | 0.708 | F=.84, df=1,<br>p=.359 |
| 1986 | 3.408<br>(N=2292) | 0.688 | F=15.53,<br>df=1,<br>p=.0001* |
| 1991 | 3.399<br>(N=2296) | 0.682 | F=10.13,<br>df=1,<br>p=.001* |
| Multivariate Test | | Univariate Test | |
| Hotelling | | F=7.97, df=3, | |

| | |
|---|---|
| Trace=0.00929<br>F=6.955,<br>df=3,<br>p=.0001*<br>(N=2249) | p=.0001,*<br>Greenhouse-Geisser<br>I=.9563 (2)<br>(N=2249) |

1 The last column indicates the contrasts which represent the difference of means in 1977 with subsequent time periods.
2. The assumption of sphericity is tenable.
* Significant at .001 level.

## Alumni Perceptions of Academic Quality

Is the favorable disposition toward one's alma mater the result of the college's contribution to the intellectual development of the alumnus? Two indexes were created to gauge students' rating of the educational institution.

The first index consists of five items asking students the extent to which they were challenged by their program, the variety of course offerings, the quality of instruction, the usefulness of the program, and the satisfaction with the Program of Study. Cronbach's alpha was computed on these five sets of items for the four time periods, and the index entitled "program satisfaction" was constructed. The program satisfaction index score for 1977, 1981, 1986, and 1991 ranged from 4 to 25. Those who were dissatisfied with the quality of academic program scored low on the scale, and those who were satisfied were on the higher end of the continuum. Cronbach's alpha and the means for all four time periods for the scale constructed is provided in Table 4. The high coefficient associated with Cronbach's alpha for all four years indicates that the items can be reliably summed up to construct a scale to measure program satisfaction (Table 4 ).

Table 4

RELIABILITY MEASURE FOR PROGRAM SATISFACTION INDEX

| Variables (1) | Mean | Standard Deviation |
|---|---|---|
| Challenged by your program of study(1977) | 3.920 | 0.960 |
| Program provided a well integrated set of courses (1977) | 3.660 | 1.021 |
| Quality of instruction in major department (1977) | 3.768 | 0.943 |
| Program of study was worthwhile (1977) | 4.020 | 0.960 |
| Satisfaction with your major program (1977) | 3.869 | 0.902 |
| Cronbach's Alpha (1977) = 0.837, (N=2264)* | 19.16 | 3.77 |
| Challenged by your program of study (1981) | 3.977 | 0.923 |
| Program provided a well | 3.758 | 0.983 |

| | | |
|---|---|---|
| integrated set of courses (1981) | | |
| Quality of instruction in major department (1981) | 3.872 | 0.887 |
| Program of study was worthwhile (1981) | 4.000 | 0.973 |
| Satisfaction with your major program (1981) | 3.883 | 0.895 |
| Cronbach's Alpha (1981) = 0.840, (N=2279)* | 19.44 | 3.67 |
| Challenged by your program of study (1986) | 4.046 | 0.896 |
| Program provided a well integrated set of courses (1986) | 3.833 | 0.942 |
| Quality of instruction in major department (1986) | 3.910 | 0.859 |
| Program of study was worthwhile (1986) | 4.037 | 0.907 |
| Satisfaction with your major program (1986) | 3.932 | 0.870 |
| Cronbach's Alpha (1986) = 0.875, (N=2285)* | 19.72 | 3.70 |
| Challenged by your program of study (1991) | 4.253 | 0.800 |
| Program provided a well integrated set of courses (1991) | 3.950 | 0.882 |
| Quality of instruction in major department (1991) | 3.980 | 0.808 |
| Program of study was worthwhile (1991) | 4.038 | 0.881 |
| Satisfaction with your major program (1991) | 3.948 | 0.847 |
| Cronbach's Alpha (1991) = 0.866, (N=2283)* | 20.13 | 3.42 |

1 Item scale ranged from 1 to 5, i.e., "low satisfaction" to "high satisfaction."

* Items were summed up to construct program satisfaction index.

The second index is called "quality of faculty guidance", and consists of three items asking students to rate the quality of academic guidance, vocational advice and the extent of communication between faculty and students regarding student needs, concerns and suggestions. Cronbach's alpha was computed on these three items for the four time periods. The faculty guidance scale for the four time periods ranged from 1 to 15. Respondents who thought that intellectual guidance was unsatisfactory were on the lower end of the spectrum and those who rated it highly were on the higher end of the scale. Cronbach's alpha and the means for all of the four time periods is provided in Table 5. The reliability coefficient was very high for these three items and the items were summed up to construct the scale.

Table 5

RELIABILITY MEASURE FOR FACULTY GUIDANCE INDEX

| Variables (1) | Mean | Standard Deviation |
|---|---|---|
| Quality of academic guidance (1977) | 3.154 | 1.215 |
| Quality of vocational guidance (1977) | 3.744 | 1.238 |
| Channels of communication between faculty and students regarding student needs, concerns and suggestions (1977) | 3.217 | 1.107 |
| Cronbach's Alpha (1977) = 0.803, (N=2234)* | 9.03 | 3.04 |
| Quality of academic guidance (1981) | 3.210 | 1.175 |
| Quality of vocational guidance (1981) | 2.720 | 1.187 |
| Channels of communication between faculty and students regarding student needs, concerns and suggestions (1981) | 3.266 | 1.052 |
| Cronbach's Alpha (1981) = 0.828, (N=2253)* | 9.14 | 2.97 |
| Quality of academic guidance (1986) | 3.243 | 1.107 |
| Quality of vocational guidance (1986) | 3.805 | 1.149 |
| Channels of communication between faculty and students regarding student needs, concerns and suggestions (1986) | 3.270 | 1.040 |
| Cronbach's Alpha (1986) = 0.841, (N=2241)* | 9.24 | 2.90 |
| Quality of academic guidance (1991) | 3.138 | 1.110 |
| Quality of vocational guidance (1991) | 2.741 | 1.200 |

| Predictors | Standadrdized Estimate | Standard Error of Beta | T value | Significance Level |
|---|---|---|---|---|
| Program satisfaction index | | | | |
| 1977 | 0.347 | 0.004 | 15.63 | 0.0001 |
| 1981 | 0.350 | 0.003 | 15.03 | 0.0001 |
| 1986 | 0.369 | 0.003 | 16.16 | 0.0001 |
| 1991 | 0.396 | 0.003 | 16.93 | 0.0001 |
| Faculty guidance Index | | | | |
| 1977 | 0.170 | 0.004 | 7.59 | 0.0001 |
| 1981 | 0.098 | 0.004 | 4.02 | 0.0001 |
| 1986 | 0.099 | 0.004 | 4.23 | 0.0001 |
| 1991 | 0.096 | 0.006 | 4.04 | 0.0001 |
| Campus (1=Urbana) | | | | |
| 1977 | 0.196 | 0.030 | 10.26 | 0.0001 |
| 1981 | 0.201 | 0.028 | 9.68 | 0.0001 |
| 1986 | 0.174 | 0.027 | 8.85 | 0.0001 |
| 1991 | 0.146 | 0.028 | 7.56 | 0.0001 |
| Bachelors (1=Bachelors) | | | | |
| 1977 | 0.104 | 0.028 | 4.71 | 0.0001 |
| 1981 | 0.208 | 0.053 | 4.49 | 0.0001 |
| 1986 | 0.136 | 0.039 | 3.79 | 0.0002 |
| 1991 | 0.146 | 0.034 | 4.92 | 0.0001 |
| Salary | | | | |
| 1981 | 0.082 | 0.000 | 3.54 | 0.0004 |
| 1991 | 0.076 | 0.000 | 3.66 | 0.0003 |
| Gender (1=Male) | | | | |
| 1977 | -0.069 | 0.028 | -3.15 | 0.001 |

| | | | |
|---|---|---|---|
| Adjusted R2=0.264<br>(1977)<br>(N=2116) | Adjusted R2=0.227<br>(1981)<br>(N=2115) | AdjustedR2=0.217<br>(1986)<br>(N=2119) | Adjusted R2=0.233<br>(1991)<br>(N=2122) |

| | | |
|---|---|---|
| Channels of communication between faculty and students regarding student needs, concerns and suggestions (1991) | 3.254 | 1.040 |
| Cronbach's Alpha (1991) = 0.836, (N=2232)* | 9.05 | 2.86 |

1 Item scale ranged from 1 to 5, i.e., "low satisfaction" to "high satisfaction".

* Items were summed up to construct faculty guidance index.

## Impact of Faculty Excellence and Program Satisfaction on Attitude Toward the University

In this section of the article, we use the two indexes as predictors to explain students' attitude towards the alma mater (See Note 2). The attitude towards the University for the four time periods was regressed on a set of demographic variables and the two indexes, and the results are displayed in Table 6. Although it makes stringent demands on the data, OLS regression estimates the collective capability of a set of independent variables to predict the values of a dependent variable, and indicates the relative predictive power of one factor net of other predictor effects. Included in the model were gender, age, degree received, campus site (Note 3), geographical location, employment status, salary earned, and the two indexes related to program satisfaction and faculty excellence. Age, salary earned and the two indexes related to program satisfaction and faculty excellence were interval scale variables and the other five predictors were coded as dichotomous (Note 4).

Table 6 reports the standardized regression estimate and standard error for each significant predictor, the critical value for each as estimated by a one-tailed T-test, the overall adjusted R2 , and the number of cases on which the model is estimated. The p values that are given in the last column of Table 6 represent the significance of each predictor in explaining the overall model. To be conservative in our estimate, the decision was made to judge the strength of each predictor at the critical value of .0015.

An inspection of data in Table 6 demonstrates that in all four waves, baccalaureate degree holders, campus location and the two scales related to program satisfaction and faculty guidance emerged as significant predictors of attitude towards the University. The data depicts that in all four waves, baccalaureates had a more positive outlook than the professionals in their attitude towards the University. In other words, one year after graduation, women baccalaureates from the Urbana campus who scored high on the program satisfaction and faculty guidance indexes had a more positive attitude toward the University. However, gender appeared as a significant variable in predicting attitude towards the University only one year after graduation. The pattern which emerges after ten years revealed that bachelor degree holders from the Urbana campus who scored high ratings on the program satisfaction and faculty guidance indexes proclaim positive feelings towards their educational institution. Interestingly, salary emerged as a significant predictor after an interval of five and fifteen years in predicting positive attitude toward the university. The data seems to indicate that satisfaction with the university is correlated with the success of baccalaureate graduates in their transition to work. How well does the first model fit the data? The overall adjusted R2 indicates a moderate fit. Measurement error undoubtedly sapped predictive potency. However, the data provides good information on factors that shape and mold attitude towards the educational institution.

Table 6

OLS REGRESSION OF ATTITUDE TOWARDS THE UNIVERSITY IN FOUR TIME PERIODS

## Conclusion

Alumni surveys have been used by colleges and universities for a number of years and for a variety of reasons. This article is a penetrating study of alumni attitudes towards the University of Illinois over a period of fifteen years. The extended period involved in this analysis helped us to appreciate the enduring influence of higher education in students' lives and the important role of a good university education. This panel data on alumni attitudes towards the academic environment indicates that contrary to evidence from past research, students develop a stronger attachment towards the educational institution with the passage of time. A similar positive pattern was evident when examining the attitude towards program major. It is possible that better experience in the real world has made the alumni evaluate the quality of education they received at the University of Illinois. Also, favorable disposition toward one's institution seems to be, to a very considerable extent, the result of the college's contribution to the intellectual development of the student. This fact was reinforced by students' high ratings on the "program satisfaction" and "faculty guidance" indexes in predicting a positive attitude toward the university.

It is evident from this analysis that the focus of colleges and universities should be on efforts to improve the quality of education through academic advising, mentoring programs and career exploration, and planning. Notably, follow up studies of graduates' employment experiences, and satisfaction with the institution and major program of study would provide valuable feedback to the University to help assess and monitor student and institution performance. Systematic graduate follow-up survey information helps set the stage for universities to review programs within different disciplines. The information obtained from the alumni survey can be used as a standard against which the university can compare the employment and satisfaction of its graduates in order to identify programs for additional review and for making program improvements. In addition, the universities can use the follow-up information in assisting currently enrolled students in program selection and career planning. At both campus and state levels, systematic information on the employment, further education, and satisfaction of graduates is important to documenting educational accountability.

It is important to study college graduates to understand the evaluation of their own educational experiences and how they envision higher education as a major social institution. Alumni research, along with other outcome measures, can be used for a variety of purposes. Applications include academic program review and evaluation, student retention, institutional planning, marketing, and public relations. Alumni outcomes can be used for assessing the effectiveness of the general education program. Information on student outcomes can be used in institutional planning and budget review at several levels. The insights derived from these surveys on students progress could be provided to employers and public on how well educational programs address labor market needs. For administrators, alumni information provides guidance about the strengths and weaknesses of various aspects of the whole university. In a broader perspective, this research has great relevance to the University's image, which affects future development in terms of public relations and student recruitment. The results of this study were intended to assist universities in program reviews and in providing a basis for improving graduates' educational experiences.

## Appendix

Repeated measures analysis is a powerful statistical design, since the variability due to individual differences is removed from the error term which causes error variances (Stevens, 1986). The three assumptions for a single group Univariate repeated measures analysis are:

- independence of observations
- multivariate normality
- sphericity

All of the above assumptions were met in our analysis. The independence of observation is by far the most important assumption, for even a small violation of it produces a substantial effect on both the level of significance and power of the F statistics (Stevens, 1986). It has been argued by some scholars that under certain conditions, independence of observations may or may not be tenable (Glass & Hopkins, 1984, p. 353):

> Whenever the treatment is individually administered, observations are independent. But where treatments involve interaction among persons, such as "discussion" method or group counseling, the observations may influence each other.

In our case, the implementation of survey questionnaire excludes any possibility of dependence among the observations.

The sphericity assumption requires that variances of the differences for all pairs of variables be equal (Stevens, 1986). In other words, the sphericity assumption states that the covariance matrix for the difference variables is a diagonal matrix, with equal variances on the diagonal. The extent to which the covariance matrix deviates from sphericity is reflected in a parameter called I (epsilon), and if sphericity is met, then I=1. The assumption of sphericity was tenable in our two repeated measures design.

Also, repeated measures analysis of variance is fairly robust (Note 6) against violation of multivariate normality. A scholar notes that "even for distributions which depart markedly from normality, sums of 50 or more observations approximate to normality" (Bock, 1975, p. 25). In our analysis, the first repeated measures design was based on 2290 observations and the second analysis had 2249 observations.

### Notes

1. There are some limitations in panel research like panel mortality, contamination through repeated measurements, and the changing meanings of instrument items (Markus, 1979). Since the research relies on data collected through a mail survey, the length of the instrument becomes a matter of concern. This constraint makes it difficult for the researcher to ask respondents all the questions one wishes to ask, e.g., those related to the life-experiences of alumni after graduation.
2. Attitude towards the University was a close-ended scale which ranged from 1 to 4, from "strongly negative" to "strongly positive."
3. The University of Illinois has two campuses at Chicago and Urbana-Champaign. The overall quality of the University places it among the nation's top institutions of higher education. However, the Urbana campus ranks much higher in terms of academic achievement than Chicago.
4. Age and Salary were coded as an open-ended scale. The two indexes related to program satisfaction and faculty guidance were created after computing Cronbach's Alpha, and then summing up the relevant items. Gender, campus, geographical location and employment status were coded as dichotomous variables, 0 or 1. The value of 1 for gender represents male students. The Urbana-Champaign campus was coded as 1. Respondents from Illinois were coded as 1 for the geographical location variable, and people who were currently employed were coded 1 for employment status. For the degree level, we created three dummy variables, Bachelors, Masters and Doctoral, and the Professional degree holders were treated as the reference group.
5. The model is being tested at a tighter alpha level to control for positive bias and to prevent any occurrence for capitalizing on chance.
6. Robust means that the actual alpha is close to the nominal alpha.

## References

Bock, R.D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw Hill.

Glass, G.V. & Hopkins, K. (1984). Statistical Methods in Education and Psychology.

Engelwood Cliffs: Prentice Hall.

Markus, Gregory B. (199). Analyzing Panel Data. Beverly Hills / London: SAGE Publications.

Melchiori, G. (1988). Alumni Research: An Introduction. New Directions for Institutional Research, 60, 1-11.

Moden, G.O. & .Williford, M.A. (1988). Applying Alumni Research to Decision Making. New Directions for Institutional Research, 60, 67-75.

Pace, C. R. (1979). Measuring Outcomes of College: Fifty Years of Findings and Recommendations for the Future. San Francisco: Jossey-Bass.

Rich, H.E. & Jolicoeur, P.H. (1978). Student Attitudes and Academic Environments: A Study of California Higher Education. New York: Praeger Publishers.

Spaeth, J.L. & Greeley, A.M. (1970). Recent Alumni and Higher Education: A Survey of College Graduates. New York: McGraw Hill Book Company.

Stevens, J. (1986). Applied Multivariate Statistics for the Social Sciences. First edition. New Jersey: Lawrence Erlbaum Associates.

Stover, W.S. (1930). Alumni Stimulation By the American College President. New York: Teachers College, Columbia University.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 7** **April 4, 1996** **ISSN 1068-2341**

---

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS
ARCHIVES.Permission is hereby granted to copy any
article provided that EDU POLICY ANALYSIS
ARCHIVES is credited and copies are not sold.

---

# Respecting the Evidence: The Achievment Crisis Remains Real

**Lawrence C. Stedman**
**State University of New York-Binghamton**

*stedman@binghamton.edu*

**Abstract:** Wherein Stedman answers Berliner and Biddle's reply to his review of *The Manufactured Crisis*.

> "It ain't so much the things we don't know that get
> us into trouble. It's the things we know that just
> ain't so." .....Artemus Ward

In his engaging book, HOW WE KNOW WHAT ISN'T SO, the social psychologist Thomas Gilovich offers marvelous insights into the origins of human misconceptions. The problem, he finds, is not irrationality but flawed rationality--the very reasoning mechanisms that help us make sense of reality also lead to questionable beliefs. These include

- the tendency to seek confirmatory information,
- the excessive impact of confirmatory information, and
- the tendency to evaluate evidence in a biased manner.

He explains that "We humans seem to be extremely good at generating ideas, theories, and explanations that have the ring of plausibility. We may be relatively deficient, however, in evaluating and testing our ideas once they are formed" (p. 59).

In a fascinating insight, he notes that people "place a premium on being rational and cognitively consistent" and so rather than simply disregard evidence, they "subtly and carefully 'massage' the evidence to make it consistent with their expectations" (p. 53).

This leads to the illusion of objectivity:

> Although people consider their beliefs to be closely tied to relevant evidence, they are generally unaware that the same evidence could be looked at differently, or that there is other, equally pertinent evidence to consider (p. 80).

One fundamental mechanism that gets us into particular trouble is what Gilovich calls "optional stopping":

> When the initial evidence supports our preferences, we are generally satisfied and terminate our search; when the initial evidence is hostile, however, we often dig deeper, hoping to find more comforting information, or to uncover reasons to believe that the original evidence was flawed (p. 82).

Or, as he puts it more directly:

> I have argued that people often resist the challenge of information that is inconsistent with their beliefs not by ignoring it, but by subjecting it to particularly intense scrutiny (p. 62).

For complex issues, such as the condition of U.S. education and achievement, the desire for consistency outweighs the willingness to respect ambiguity.

> For nearly all complex issues, the evidence is fraught with ambiguity and open to alternative interpretation. One way that our desires or preferences serve to resolve these ambiguities in our favor is by keeping our investigative engines running until we uncover information that permits a conclusion that we find comforting (p. 83).

Gilovich has captured well the fundamental failing of the MANUFACTURED CRISIS. Whether Berliner and Biddle are discussing the "myths" about achievement and schools, the power of right- wing disinformation, or the contrast between neoconservative and progressive reforms, they repeatedly offer a one-sided treatment of the evidence. With few exceptions, they accept at face value any information that supports their viewpoint, while they dissect and reinterpret any information that challenges it.

The purpose of academic training and scholarship is to rise above such flawed rationality; to learn how to critically analyze the evidence that supports your own favored arguments--and to treat fairly the evidence that contradicts it. It is also a matter of learning to accept the complexity and ambiguity of evidence--and to fairly present that.

Unfortunately, Berliner and Biddle failed to do this--either in their book or in their response to me. They have even gone beyond the flawed rationality Gilovich describes. They ignored or dismissed entire areas of relevant evidence--such as the extensive data on students' low levels of achievement and knowledge--and, in selectively presenting other evidence--such as the data on test score trends--they winnowed out only that which supported their viewpoint and discarded the rest. In several cases, they have even directly misrepresented the actual data.

What's worse is that they are now resorting to sweeping, disrespectful condemnations of those who disagree with their arguments and point out the limitations of their evidence. They characterize the various critiques of their book as "distorted portrayals and outright lies"; they labeled my analysis a "diatribe" and as "disingenuous" and filled with "lacunae, misrepresentations, and trivialities". They have impugned the motives of reviewers and, in my case, even attributed positions to me that I have never taken.

This, too, is understandable, however. As Gilovich points out, the psychologist Robert Abelson argued that "beliefs are like possessions" and that, consequently, people are "possessive and protective" of them and react defensively when their limitations are pointed out (pp. 86-87). The motivational determinants of belief are particularly powerful. As Sir Francis Bacon put it in the NOVUM ORGANUM, "Man prefers to believe what he prefers to be true" (Gilovich, 1991, p. 75).

## THE PURPOSE OF MY EPAA REVIEW

Berliner and Biddle were upset that my review focused on their treatment of the achievement evidence. That was its purpose and should have been immediately obvious from the introductory paragraphs.

Why did I focus on the achievement evidence? Because it underpins their basic argument about a manufactured crisis. They claim that U.S. students and schools are actually doing well and that the evidence to the contrary and beliefs in a crisis have been manufactured by right-wing school critics and administrations. Having already produced a general review of their book back in November in the WASHINGTON POST (one I am sure they must have seen) (Stedman, 1995), I felt it imperative to discuss, at length, in an academic forum, the details of how they treated the evidence on student achievement. EDUCATION WEEK also had devoted a full-page general story about their book back in September (Viadero, 1995).

It should be noted at the outset that even Berliner and Biddle considered such evidence so central to their argument that they spent several chapters trying to explode the myths about the current condition of schools and achievement.

Contrary to their repeated claim in their response, I never stated that "their book" was based on four sweeping claims, but rather that their achievement analysis was. Nevertheless, the review was supposed to have contained the following two introductory sentences, which could have eliminated much of their consternation.

> This review is focused on the achievement analysis portion of the MANUFACTURED CRISIS. My more general review of the book can be found in the Education Review section of the Washington Post, Sunday, November 5, 1995, pages 16-17.

## OVERVIEW: THE MAJOR FAILINGS OF THEIR ACHIEVEMENT ANALYSIS

The actual evidence on student achievement is crucial to their argument. It directly addresses their claim that U.S. students are achieving well and that the educational crisis has been "manufactured". Instead of systematically reviewing the evidence, they selected a few pieces of data on each topic and reinterpreted them to suit their argument. They concentrated on trends (mostly stable) but ignored levels of achievement (mostly low).

Let me be clear at the outset. I believe that right-wing forces have been attacking the public schools and EXPLOITING the evidence, but there is also extensive, credible evidence that there is a real achievement crisis, something Berliner and Biddle continue to deny. They have still never dealt directly with the actual evidence about low achievement.

Their response to my review repeats and reinforces the book's major failings in its treatment of the achievement evidence. Here's what they did (or did not do) in their analysis.

1. They ignored a large and growing body of research which shows that student achievement has been weak for several decades. Our high school students lack important knowledge in history, civics, geography, and English; they have done poorly in mathematics and science and few write well. The evidence is overwhelming that the achievement crisis is real. In the next section, I report on the latest National Assessment of Educational Progress (NAEP) results.
2. They analyzed the test score decline in a misleading fashion. Although they rightfully criticized the myth of a RECENT general achievement decline, they ignored the 1970s decline and failed to present any of the contradictory evidence from the 1980s. They clearly overstated the case when they claimed "only ONE test, the SAT" ever suggested a decline (p. 35 emphasis original).

   Worse, they then overreached and tried to cast current achievement in an historically positive light. Without the needed evidence, they claimed that this generation of students achieves "substantially" higher than previous ones on "virtually all" commercial standardized tests--a contention that is directly refuted by the major reviews of historical trends on such tests.

   In their response, they compounded their error by arguing that then-and-now studies--including MY review of such research--support such a sweeping contention. They claimed that "almost all" then-and-now studies showed improvement when, in fact, many studies showed no change, several showed declines, and the ones showing improvement typically involved small gains (Stedman & Kaestle, 1991b). They did not mention that such studies have been fraught with problems. They also have never acknowledged that achievement on NAEP HIGH SCHOOL science and civics tests remains lower than in the past, below their 1969 levels.

3. They tried to claim that U.S. failure in international assessments is a "myth", but it is actually partly true. Although our younger students have done well in reading, our older students have done quite poorly in secondary school math and the high school sciences. Here, again, they overreached by claiming that U.S. schools "stack up very well" in the international comparisons. For one thing, they argued that curriculum differences were a major cause of the international achievement differences, but they based this on only one study of outdated 8th grade math data from 1981- 82, and the data did not support their claim. This was a thin reed on which to characterize the standing of the U.S., and even if it had been true, it is still disturbing news for it means the U.S. curricula and programs are not up to international standards. More recent studies also do not support their assertions. In the 1991 IAEP math study, our 8th graders lagged well behind those in nearly all other countries, and this was true even when algebra curricular differences were accounted for.
4. They systematically misrepresented major research studies and data on U.S. achievement.

   a) They graphed standardized test score trends from a study by Linn, Graue, and Sanders (1990), but somehow dropped the very tests and grade levels which included declines! Worse, they offered these data as definitive proof of improving achievement, when in fact, Linn, Graue, and Sanders pointedly remarked that the results were "equivocal" and noted that part of the gains were caused by districts' repeated use of the same tests rather than by genuine improvement. The 1980s back-to- basics movement also helped to artificially raise scores by frequent testing and skill-drill approaches (Stedman & Kaestle, 1991a). In their response, they claimed that the omitted data supports their original claims when, in fact, much of it contradicts them. The data were also outdated, coming from the late 1970s through mid-1980s, and thus are not even relevant to their claims about current students or recent improvement. In a later section, I will discuss their continued mischaracterizations of this study and its data.

   b) They graphed international math scores from a study by Westbury (1992), but somehow left out his 12th grade comparison where the U.S. did poorly! Worse, they disregarded Westbury's caution and improperly compared our elite 8th grade algebra students to the AVERAGE Japanese student. Westbury actually used the top 20%. They claimed it proved that with a COMPARABLE curriculum our students do well in math, but never mentioned that our students spent far more time on algebra (61% vs. 26%), covered more test items, and were one grade older.

   c) They claimed the international assessments have improperly compared the broad mass of U.S. students to an overseas elite attending high-status high schools, but this is old criticism from the early international studies, and it was only partly true even back then. In the early IEA math studies, for example, researchers deliberately sampled college-bound students who were taking math in their senior year of high school--in the U.S. this was an elite group of only 18% of our students; in the second IEA math study, it was only 13%, a similar percentage to that in other countries (Stedman, 1994a).

   d) They attributed the SAT decline to demographic changes in test takers, yet never reviewed the evidence which shows this explains much, but not all, of the decline. They also used AVERAGE SAT scores to claim minority student performance gains, but this masked minority VERBAL declines in the late 1970s and late 1980s.

These are serious, major failings (not molehills) which directly undermine their argument and impugn their credibility as scholars. It is little wonder that they chose to attack me personally rather than deal forthrightly with the evidence.

I have divided this response into several sections:

> THE NEW ACHIEVEMENT EVIDENCE--a review of the 1994 NAEP findings which shows that students continue to display serious weaknesses in their knowledge and skills.
>
> THE EVIDENCE AND THEIR RESPONSE--a direct response to their arguments in their reply, organized around their four sweeping claims about U.S. achievement which they continue to support.
>
> THE MANUFACTURED CRISIS REVISITED--a look at several major areas of errors and misrepresentation that were not covered in my original review, in particular their claims of high levels of parental satisfaction with local schools. Here again, they were so intent on fitting the data to their argument, that they distorted the evidence. It turns out that only about a quarter of public school parents rate their oldest child's school an A, while about half of them rate their community's schools C through Fail!
>
> PROGRESSIVE REFORMS AND THE RIGHT-WING AGENDA--an endorsement of much of their reform agenda, coupled with an analysis of their one-sided presentation of a national right- wing agenda, which again demonstrates their Procrustean handling of evidence. In particular, I discuss their treatment of the Sandia Report, which they claimed provided a valid look at the achievement evidence and which they allege was suppressed by the Bush administration.

## THE NEW ACHIEVEMENT EVIDENCE

Students are struggling. The depth of the achievement problem is strongly borne out by the latest round of NAEP studies of reading, history, and geography achievement. Performance is reported for basic, proficient, and advanced levels. In 1994, substantial portions of students did not even make the basic level while a majority failed to achieve the proficient level in each subject at each grade level tested: 4th, 8th, and 12th. I also review the results from NAEP's 1992 assessment of writing portfolios, which revealed that little classroom writing is of high quality.

## 1994 HIGH SCHOOL SENIORS' ACHIEVEMENT

I concentrate here on the data for high school seniors because they provide the best overall assessment of K-12 performance. In reading, a quarter of our seniors failed to reach even the basic level (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995, p. 15). Only about one-third demonstrated reading proficiency (or better). In geography, about a third were below the basic level, while only about a quarter displayed proficiency (or better) (Williams, Reese, Lazer, & Shakrani, 1995, p. 16). History showed the worst results. Over half the seniors were below the basic level and only 11% made the proficient level or higher (Williams, Lazer, Reese, & Carr, 1995, p. 19).

These levels were set by NAEP's independent policy-making body-- the National Assessment Governing Board with "contributions from a wide variety of educators, business and government leaders, and interested citizens" (Williams, Reese, Lazer, & Shakrani, 1995, p. 3).

> The reader should recognize that the results are based on the judgments of panels, approved by the Governing Board, of what advanced, proficient, and basic students should know and be able to do in each subject assessed (p. 9).

Concerns have been raised about the construction and interpretation of these levels

(Stedman, 1993), and this latest series of NAEP report cards clearly labels them as "developmental" (Williams, Reese, Lazer, & Shakrani, 1995, p. 3). Nevertheless, both the Commissioner of the National Center for Education Statistics and the National Assessment Governing Board believe the levels are "useful and valuable" in reporting on student achievement.

Fortunately, NAEP has returned to their practice of making public sets of test items used in the assessments. This allows educators and the public to appraise the items and evaluate student knowledge directly. The test themselves are quite rich, combining constructed response questions with multiple choice ones. In geography, for example, 60% of the testing time was devoted to constructed response items. The geography and history tests offer a rich panoply of maps, graphs, photographs, cartoons, paintings, and magazine covers. A look at individual items avoids the scaling problems and reveals that many students have serious deficiencies in basic knowledge and skills.

## 1994 GEOGRAPHY RESULTS

Let's consider the geography results first (Williams, Reese, Lazer, & Shakrani, 1995). Less than half the seniors knew that slavery was a major reason many Caribbean people are of West African descent (p. 63). Only about a third recognized a description of a rain forest and could identify a country that had one. Only about a quarter could identify three or more of the following on a map--the Pyrenees Mountains, the Japanese Archipelago, the Mediterranean Sea, and the Persian Gulf. (And this was after the Persian Gulf War!) Only 10% could interpret a simple bar chart of predicted hydrocarbon emissions and give a reason for the trends displayed.

Relatively stronger results were found for identifying four world cities as major religious centers (76%), identifying shaded countries on a world map as belonging to OPEC (65%), and deciphering interpreting tabular data about two countries (53%- 67%). Still, it should be noted that one-fourth to over one- third of the students had problems with such items.

## 1994 HISTORY RESULTS

In history, the results were also disturbing (Williams, Lazer, Reese, & Carr, 1995). Only about half of the high school seniors (55%) knew that cotton trade was a main reason Great Britain leaned toward the Confederacy during the Civil War. The other choices were British plantation owners held slaves, most British immigrants lived in the South, and British politicians wanted to conquer the U.S.

Less than half of seniors could identify the purpose of the Monroe Doctrine (41%), date a newspaper report about the Civil War destruction of Charleston (41%), or realized that preventing the spread of communism dominated U.S. foreign policy in the post- war period (47%).

Less than half (47%) could interpret an 1876 magazine cover depicting the "Indian problem" even though general statements were permitted about attitudes or events. Only a third were able to identify a consequence of Nat Turner's slave rebellion (tighter controls on slaves). Only a quarter know that the Camp David accords promoted peace between Egypt and the U.S. (Other choices were the Soviet Union and China; Palestinians and Jordanians; North Korea and the U.S.). Only 15% were able to interpret a simple cartoon showing the long, winding road necessary to spiritually fulfill the civil rights law after enactment.

There were several strong spots. Over 80% properly interpreted two paintings of George Washington as reflecting the glorification of political figures and the use of religious symbols and, in what was hardly a surprising result, 88% knew that the computer rather than the typewriter, superconductor, or radio produced the greatest change in how people worked between 1960 and 1990.

## OTHER RECENT NAEP EVIDENCE ABOUT STUDENT PERFORMANCE

Writing is another area that is important particularly given the connection between critical thinking skills and written expression. In 1992, NAEP conducted the first national assessment of writing PORTFOLIOS gathered from classrooms across the country. Such an approach avoids the artificiality and time pressures of using a national sit-down test to judge writing ability. The findings were troubling. Olson (1995) reported that "the best writing that students produce as part of their classroom work is still not very good."

Only between 4-12% of the 8th graders achieved high marks (5 or 6) on the six-category evaluation scale. One-fourth to almost one-half received low marks (1 or 2), depending on whether informative or narrative tasks were being considered. Gary W. Phillips, the associate commissioner at the National Center for Education Statistics, concluded that, "The moral of the story is that the writing is not very good in the nation. Even the best is mediocre."

This may be a bit harsh, however, given that there were writing samples achieving the highest ratings. The portfolio assessment methodology also needs to be systematically and independently evaluated. No doubt, problems will be found that could require some adjustments to the results (up or down). In the meantime, though, the findings suggest there is a serious writing problem and mirror those of the traditional set-task writing assessments that NAEP has conducted, including the one in 1992 (Applebee et al., 1994). High school students have struggled over the years. In 1992, only about 2% to 23% produced "elaborated or better" writing, with the weakest performance on persuasive tasks (Applebee et al., 1994, p. 5). (These are averages across four tasks of each type: persuasive, narrative, and informative. On one informative task, students did much better, 46%; on another much poorer, 6%.) The percentages who produced "developed or better" responses was better but still troubling--only around 16% to half of the students performed acceptably. On most tasks, most students' writing was undeveloped or minimally developed. This mirrors their inadequate writing in prior NAEP assessements (Applebee et al., 1990, p.107; see Stedman, 1993 for information about earlier results and scoring methods.) The good news is that most students have done well with basic mechanics--spelling, grammar, and punctuation--and so additional WHOLE-class drill and practice in these areas is not warranted.

NAEP also did a follow-up analysis of the 1992 reading assessment in which they explored student performance on different kinds of test questions (Olson, 1995). They found a marked drop-off in student understanding and proficiency as the questions became more open-ended and required more elaborated responses. At the three grade levels (4, 8, and 12), performance fell from around two-thirds correct on multiple-choice problems, to slightly above half on short, constructed answer questions, and then to only one-fourth to around a third on questions requiring an extended response.

All of which has important implications as we move toward more authentic assessment. We will most certainly find initially that student performance is even worse than what has been revealed by the more straight-forward, multiple-choice recall testing that has been done primarily so far.

In math, NAEP analysts have determined that "less than half (of high school seniors) appeared to have a firm grasp of seventh- grade content" and only 5 percent "attained a level of performance characterized by algebra and geometry--when most have had some coursework in these subjects" (Mullis et al., 1991b, p. 80). Although high school students have done well on basic operations such as adding whole numbers and reading a line graph (90%+), many have trouble even with simple problems involving fractions, decimals, and percents (Mullis et al., 1991, pp. 302- 309). In 1990, for example, 34% of 17-year-olds could not find the area of a rectangle, given a diagram and the length of two sides (Mullis et al., 1991a, p. 306). Math educators who reviewed the NAEP data in the late 1980s determined that students "exhibit serious gaps in their knowledge and are learning a number of concepts and skills at a superficial level" (Carpenter et al., 1988, pp. 40-41). They concluded that "students' achievement at all age levels shows major deficiencies." Although there have been some modest gains in math achievement in the 1990s, their general conclusions are still appropriate today.

By the way, the NAEP findings I have presented do NOT include dropouts; overall high school student achievement is, therefore, likely to be even worse than this evidence indicates. When we combine these recent results with those from the past several decades, we have a serious cause for concern. (See Stedman, 1993, for a review of this evidence and a discussion of its strengths and limitations.)

## BERLINER & BIDDLE'S REJOINDERS

Instead of reviewing this extensive and troubling evidence about low achievement, Berliner and Biddle offered a series of rejoinders in their book about unrealistic standards, our students' focus on breadth of experience, and the nature of the tests. As I explained in my review, however, the achievement standards are realistic (they might even have been set too low), knowledge is an important part of our students' experience, the achievement problems are not an artifact of psychometric scaling, and the tests incorporated real-world tasks and knowledge.

In their response, they took much the same approach. First, they wrote that the "standards against which America's schools are to be judged and found wanting are arbitrary and can be made up as one goes along". Historically, this is untrue. The major studies have not used "arbitrary" or "made up" standards; they have relied strongly on school- and curriculum-based measures-- the textbooks that are most widely used, teacher consensus about what is important to be tested, citizen panels on what students should know and be able to do. Most people would certainly expect high school seniors to have mastered 7th grade math and basic social studies, but they have not.

Berliner and Biddle then suggested that those of us who are concerned about academic achievement are "school bashers" and "standardized test enthusiasts". (I, for one, am neither!) They label the solid evidence that U.S. general knowledge and academic achievement have been low for decades as "Nonsense!" and "ludicrous" (see Stedman, 1993 for a review of the evidence). That is the level of their argumentation--dismissive and mocking, without ever examining the actual evidence. Their primary argument about historically low achievement was the following:

> We find it ludicrous that anyone should claim that "academic and general knowledge have been at low levels for decades" in this country. If this were actually true, how on earth did our nation ever manage to win World War II, send astronauts to the moon, create a plethora of new pharmaceuticals, and invent the transistor and virtually all the computer technology now used world wide? For that matter, how did we achieve the world's highest rate of industrial productivity, and establish ourselves as this century's dominant super-power? "Low levels" of academic and general knowledge? What nonsense!

Let's examine this argument. These accomplishments did not depend upon the MASS of U.S. students and adults being well- informed and knowledgeable. Instead, they exemplify the prowess of the military-industrial complex in post-war America, the skills of a narrow technical elite, and the inventiveness of a single individual or group of individuals.

It took a Jonas Salk to develop the polio vaccine, for example. The transistor was invented by John Bardeen, Walter Brattain, and William Shockley. (This is the same Shockley who later espoused racially-charged ideas about intelligence being genetically determined.) The micro-computer revolution can be largely credited to three school dropouts--Steve Jobs and Steve Wozniak who developed the Apple II computer and Bill Gates who founded Microsoft.

In other words, such accomplishments have readily existed alongside low levels of knowledge and achievement in the general population. Our citizens' lack of knowledge of civics, history, geography, and literature, for example, had little bearing on our winning World War II or getting a man to the moon. (Let us also be careful lest we believe that it is only Americans who have discovered pharmaceuticals or that only a U.S. education was involved. Penicillin, for example, was developed by Alexander Fleming, but he was a Scottish biologist. Streptomycin was discovered by the American Selman Waksman, but he was born in Russia in 1888. The oral form of the polio vaccine was developed by the Polish-American Edward Sabin, born in 1906. Many of these discoverers, therefore, were educated well before World War II, long before the decades of low achievement that I was talking about!)

I find it curious that Berliner and Biddle have unwittingly embraced here a Human Capital view of economic productivity and military-corporate power, a view that they critique at great lengths in their book! According to their new argument, students' general knowledge and academic achievement have been the keys to U.S. economic and technical accomplishment!

In their book, however, Berliner and Biddle gave only a passing nod to the importance of knowledge and cultural heritage- -even for social and civic reasons. Yet it is important that students be well informed about the key events, people, issues, literary works, and social struggles that have shaped our multicultural society. Such information matters--it helps us as voters, workers, readers, newswatchers, and community members. In a society torn by debates over immigration and affirmative action, we all should be alarmed by how little our students know of world cultures and how poorly informed they are about our country's tortured racial history.

The low levels of achievement also are unimpressive results for 12 years of schooling. The tests do measure much of what is being taught in our schools and show we are not succeeding in our efforts. This is the heart of the achievement crisis. A complex, democratic society needs a well-read and knowledgeable citizenry and yet the evidence shows we are not accomplishing this.

### THE EVIDENCE AND BERLINER & BIDDLE'S RESPONSE

## SWEEPING CLAIM #1: "TODAY'S STUDENTS ARE OUT-ACHIEVING THEIR PARENTS SUBSTANTIALLY"

Their treatment of the achievement evidence continues to be one-sided. In their response, they wrote that "we were actually quite cautious in what we claimed about the achievements of students and their parents." That claim contrasts strikingly with what they actually stated in their book about standardized test trends. They claimed that "virtually all of them would show that today's students are out-achieving their parents substantially" (p. 33). Not some of them, but virtually all of them. Not somewhat outperforming, but substantially. As I noted in my review, they did not present the evidence needed to support this sweeping generational claim; they failed to discuss the many reviews of historical trends that refute it.

They then had the amazing chutzpah to cite my own research on then-and-now studies to try to prove their claim. Note first, that they did not cite this research in their book, but are only bringing it in now, after the fact. Next, notice what they claimed I found:

> Additionally, when one looks at more than 20 "then" and "now studies of student achievement--reviewed previously by Stedman himself in his studies of literacy in the U. S.!-- almost all the results show that the students taking the test "now" outscore the students that took the test "then."

They claim that "almost all the results" showed improvement. In fact, of the 13 local then-and-now studies done through the 1960s, seven showed no real change, including two that showed declines. Two of three then-and-now studies done in the 1970s showed declines relative to earlier students. Overall, across the century, more studies had gains than declines, but the gains were small and many trends were stable. The studies also suffered from a variety of flaws. Here's how Carl Kaestle and I (1991b) actually summarized our findings:

> If one takes age into account, more of the tests showed gains than declines, whereas many others showed approximately equal performance rates. But few of the studies were nationally representative. And the magnitude of the changes, up or down, was usually half a school year or less--a shift that can easily be attributed to the margin or error caused by the problems we have described (p. 89).

We then concluded:

> Our educated guess is that schoolchildren of the same age and socioeconomic status have been performing at similar levels throughout most of the twentieth century (we consider the 1970s in detail in Chapter 4). But we also caution that then-and-now

> studies are fraught with design and interpretation problems; reliance upon them to support arguments about literacy trends is unjustified (p. 89).

This illustrates well their treatment of evidence--a misrepresentation of findings and other scholars' research, a continued effort to fit the evidence to their argument, and a failure to acknowledge the complexity and problems with the data.

Note as well that they completely disregarded one of the major conclusions of our literacy research. By focusing on trends, they again ignored the findings about the levels or depth of the achievement and illiteracy problems. We wrote:

> Does this mean that things are rosy on the literacy front? Certainly not. The functional-literacy tests showed that a substantial portion of the population, from 20 to 30 percent, has difficulty coping with common reading tasks and materials. The job literacy measures, for all their limitations, show that there are substantial mismatches between many workers' literacy skills and the reading demands of their jobs. Even if schools are performing about as well as they have in the past, they have never excelled at educating minorities and the poor or at teaching higher-order skills (p. 128).

As I pointed out in my review, Berliner and Biddle selectively presented evidence on recent trends in commercial test scores, specifically data from a study by Linn, Graue, and Sanders. Remarkably, in presenting the data, they omitted the very grades and tests that showed declines and only graphed those that showed gains! They also never mentioned that the researchers had determined that the test increases were partly caused by districts' repeated use of the same tests rather than by genuine improvement.

Their explanation of their selectivity is a curious one--and should have been presented in their book, not after the fact now! First, as to their omission of SRA data--which showed reading and math declines in several grades--they argued that the SRA data are "complex and mixed, and we judged that they required too much explanation to warrant their inclusion in a book designed for general readers". That is both unscholarly and an insult to readers. They were, in fact, able to describe the data in a only few sentences in their response. It would have been easy for them to have included an extra bar in their graph covering the SRA data. They ponder: "What on earth would readers have gained had we displayed these data in TMC?" That is the nub of their problematic treatment of evidence. Readers would have gotten an honest and more complete look at the elementary school data. SRA reading scores, for example, declined in 5 of the 8 elementary school grades!

Their characterization of the data also varies, depending upon whether they are trying to support their case or discredit other researcher's positions. In their book (p. 31), for example, they described annual gains of 2 percentile points on commercial tests as "large"; yet in their footnote in their response, when they are trying to discount the significance of the SRA data, they labeled annual reading declines of 1.5 percentile points as "tiny"! (Note as well that half the "gains" they did graph were under 1.5 points!)

Second, as to their omission of high school data--which also showed some declines and where gains were less impressive--they now explain that they omitted them because high school students show less growth in academic subjects--yet wasn't that worth presenting?--and that Linn, Graue, and Sanders did not include CTBS and ITBS high school data. This is a weak excuse. A scholar interested in presenting a thorough picture would have gotten the CTBS high school data, while ITBS doesn't even go the high school level! (Riverside Publishing uses the ITED for high school students.) Furthermore, why not present the data that was at hand?

The difficulty may have been that results would not have fit their thesis as well. On the MAT, reading scores were up in 9th grade, but they declined in grades 10, 11, and 12. On the SRA, grades 11 and 12 showed declines in both reading and math. Overall, the CAT and Stanford showed annual gains of only around 1 percentile point in reading and math, much less than the elementary school scores. Given such mixed evidence, it is misleading, therefore, for them to claim that the "high school data SUPPORT our assertions" (emphasis original!).

Their characterizations of specific high school data were also questionable--as to the

MAT math scores, they wrote "ALL four high school grades provided evidence of increased scores in mathematics" (emphasis original) when in fact, 9th graders showed no change! As to MAT reading scores, they wrote, "The MAT reading tests generated mixed data for these four grades: scores were up in two grades, but scores were down in two others". As we have noted, however, scores in the last three grades 10-12 actually declined, by -.7, -.4, and -.7. Such repeated errors lead one to distrust their analysis. It should be noted, as well, that they never informed the reader that they were graphing only elementary school data--instead, they presented it as if were generally representative of student achievement, when it was not.

Finally, they still have not acknowledged that K-8 test score increases should not be simply equated with improvement in achievement. The Lake WoeBeGone phenomenon of repeated test administrations and teaching-to-the test is too well- established to be ignored. Furthermore, the 1980s back-to-basics movement also helped to artificially raise students' scores by emphasizing frequent testing and skill-drill approaches (Stedman & Kaestle, 1991a). Berliner and Biddle's conclusion, however, continues their overall sweeping characterization of the data and this study: "So, student achievement is UP on commercial tests, and that is exactly what we concluded."

One final note--this evidence is outdated, so it does not support claims about current achievement trends! The renorming data covered the period from the late 1970s through the mid 1980s. The CAT test, for example, came from a 1978-1985 renorming. The CTBS data came from a 1987 renorming. The data, therefore, is not recent, but refers to trends from over a decade ago!

## SWEEPING CLAIM #2: ONLY THE SAT EVER SHOWED A DECLINE

Berliner and Biddle were right to challenge the mythology that we are currently in a massive, general decline. We are not. But they went well beyond that in their own assertions. They wrote, "The two of us know of only ONE test, the SAT, that ever suggested such a decline" (p. 35). That is a sweeping claim and one that is unsupported by the evidence.

As I pointed out in my review, many major tests showed declines, particularly in the 1970s and at the high school level. These declines electrified portions of the legislative, educational, and public communities--they led to major investigations, including the College Board's ON FURTHER EXAMINATION (Wirtz, 1977). While conservative critics may have exaggerated their significance, the declines did occur and to claim otherwise misleads readers. Unfortunately, they did not discuss this evidence in their response--or explain their claim.

Scholars have a responsibility to present the full story, particularly contradictory evidence. Although trends have generally been stable, there are important exceptions. Berliner and Biddle never mentioned in their book that high school students' NAEP science and civics scores remain below their 1969 level, that high school reading scores fell in the late 1980s on several tests, and that the SRA tests showed reading and math declines at several grades.

Their attempt now to discredit this evidence is curious. I noted that HIGH SCHOOL students' NAEP science and civics scores had declined substantially in the 1969-1976 period. They tried to challenge this with RECENT data from 9- and 13-year olds! That was hardly relevant to my original comment. High school students' scores are also a more important indicator of performance as they reflect the entire K-12 experience.

I also noted that high school students' civics scores slipped in the late 1980s, something they took issue with. In NAEP's report, THE CIVICS REPORT CARD, however, analysts noted "Seventeen-year-olds participating in the 1988 assessment performed significantly less well than their counterparts assessed in either 1976 or 1982" (Anderson et al., 1990, p. 13).

And my judgment about science trends is not "simply wrong!" as they gleefully exclaimed. I stated that HIGH SCHOOL students' science scores "fell during the 1970s and have only partly rebounded". They even presented the data that bears me out in their response--17-year-olds had a scale score of 305 in 1969 (not 1970) and it dropped steadily to 283 by 1982--this was a substantial drop of about a half a standard deviation. By 1992, it had recovered to 294, or only about half the way back.

There was also some slippage in reading and writing scores, particularly for younger students. 9-year-olds dropped six scale points in NAEP reading achievement between 1980 and 1990 while 8th graders dropped 10 scale points in writing proficiency between 1984 and

1990. The latest reading assessment showed that 4th graders had dropped a minor three scale points between 1992 and 1994, while 12th graders had dropped five (Williams, Reese, Campbell, Mazzeo, & Phillips, 1995, p.7).

Berliner and Biddle argued that "Stedman's interpretation of the data is once again wrong! He sees a decline in reading scores when he should be seeing remarkable consistency of scores over time." This is far-fetched. I am no supporter of the decline thesis--as they well know--and stated so quite clearly in my review. In my general review of achievement trends (Stedman, 1993), which I cited in support of my comments, I wrote:

> I begin with literacy because it undergirds academic performance and is a perennial concern of educators. Here, a picture is worth a thousand words (see Figure 1). The picture for NAEP writing performance is similar to that for reading: both have remained basically stable for more than two decades (p. 216).

They also claimed that I ignored the accomplishments of schools "in the face of escalating social problems", yet in my EPAA review of their book, I wrote:

> Given changing school populations and societal conditions, generally stable scores are still a remarkable accomplishment for U.S. schools. This is an important message that the public needs to hear.

Such severe distortions and misrepresentations do them no credit.

## THE SAT DECLINE

Finally, there is the SAT decline itself. Here again, they attributed to me a position I did not take. They know I am no fan of the SAT; I have described it as an "irrelevant measure" of educational quality and national achievement (Stedman, 1994b). Others disagree, however, and so it remains of interest. Indeed, its national prominence is one reason they dealt with it. My concern again is their unscholarly and one-sided treatment of the evidence. The first problem was that they attributed the SAT decline to demographic changes in test takers, such as increases in minority students, yet never reviewed the research!

The major investigations have concluded that the SAT decline was not entirely compositional (Stedman, 1993; Stedman & Kaestle, 1991). The tremendous rise in minority test-takers, for example, cannot explain the large decline in WHITE students' SAT scores during the 1960s and 1970s. During one stretch, the pool of test takers did not expand, yet scores still declined. This suggests that, to some extent, there was a real decline in performance.

The most comprehensive analysis of the demographic changes-- the College Board's special Advisory Panel study published in 1977 (Wirtz, 1977)--concluded that much of the 1960s decline, from 2/3rds to 3/4ths, but a smaller part of the 1970s decline, up to 30%, was due to demographic changes in test takers. (They reviewed a vast array of demographic indicators.) If one considers the additional effects of age (students were getting younger) and birth order (younger siblings score more poorly), up to one-half of the 1970s decline may have been due to compositional changes. The Advisory Panel attributed the remaining portion to an UNDETERMINED combination of school and societal factors.

They may have misgivings about such research, but it was incumbent upon them to acknowledge its existence.

Curiously, in spite of their misgivings about SAT scores themselves, they chose to use them to claim that minority students gained in achievement in recent decades. They even went so far as to present a bar graph of SAT scores by minority groups to document their claim. The problem, as I pointed out, was that they used AVERAGE SAT scores which masked minority verbal declines in the late 1970s and late 1980s (Stedman, 1994b). Here again, I find it remarkable that when an error is pointed out, they do not discuss the evidence pertaining to it. Instead, they again attributed a position to me that I have never taken--that the SAT is as meaningful a barometer as NAEP. Why can they not gracefully acknowledge contradictory evidence or their errors?

(It should also be noted that they essentially set up something of a straw man argument about the decline in their book. Several of the leading conservative critics have NOT focused on the decline for some time--these educators recognize and have acknowledged that scores recently have been stable. The so-called "myth" is no longer one in certain quarters.)

## SWEEPING CLAIM #3: U.S. STUDENTS "STACK UP VERY WELL" IN INTERNATIONAL COMPARISONS

The first problem here is that the so-called "myth" of U.S. international failure is actually partly true. U.S. international performance has been dismal in secondary school mathematics and poor in several high school sciences. As I explained in my major review of the international assessments, these are real results and not an artifact caused by sampling or curricular-test bias (Stedman, 1994a). Berliner and Biddle, however, do not accept ANY evidence that shows U.S. achievement in a negative light.

The second problem is that they failed to review and summarize the findings about U.S. achievement from the major international assessments. This would have led readers to a very different conclusion about the current state of U.S. international performance. As I noted in my review, our students have "done well in reading and elementary school science, middling to poor in geography and secondary school science, and last or near-last in mathematics." That is a fair and balanced characterization of the international findings and shows that critics who make sweeping claims about a GENERAL U.S. failure are mistaken, but so are reviewers such as Berliner and Biddle who try to cast the international findings only in a positive light.

Curiously, they now write that they decided against presenting these findings because the international validity problems are so great. Yet this did not prevent them from making sweeping claims about the findings such as "Many, perhaps most, of the studies' results were generated by differences in curricula" (p. 63). A more scholarly approach, particularly for the general public, would have been to have presented the overall findings and then discussed their strengths and limitations. Nor did they present any counter-arguments or counter-evidence to their sweeping assertions about validity (I review their claims below; see also Stedman, 1994a).

The third problem is that Berliner and Biddle went well beyond challenging the mythology of a general U.S. international failure and reinterpreted selective evidence into a highly positive, one- sided view. They wrote that "American schools stack up very well" (p. 63), the international evidence "confirms impressive strengths of American education" (p. 64), and when opportunities to learn are considered, "American students' school achievement looks quite similar to that of students from other countries" (p. 58). Such sweeping contentions would not have been supportable by a general review of the international research.

### WESTBURY STUDY AND THE PRINCIPLE OF CONTROL REVISITED

One of their most egregious examples of reinterpreting evidence was their handling of Westbury's (1992) study, which was their major piece of "evidence" about curricular opportunities-to- learn. Comparing U.S. algebra students to the average Japanese student, however, violated their own research precept--the Principle of Control. As they put it,

> to estimate the true effect of a factor using survey data one MUST control, in the analysis, for the effects of other crucial factors that can affect the relationship. Trained data analysts are very aware of this principle--indeed, it one of the first things taught in courses on statistics (p. 159).

Clearly, U.S. students who take algebra in the 8th grade are a unique, elite group with marked advantages in college expectations, math interest, parental support, social class, and academic ethic. Consequently, one cannot tell how much of their achievement reflects the effects of their curriculum and how much their background advantages. The comparison is, therefore, inappropriate and unwarranted and was specifically cautioned against by Westbury himself (1992).

Furthermore, our algebra students actually had a more focused algebra program--they had spent 61% of their time on it compared to only 26% for Japanese students. They also had covered more test items and were one grade older. So even the curricula--or opportunities to learn--were not similar as Berliner and Biddle asserted. (They also labeled the data as "achievement scores" when in fact it was only algebra scores.)

In general, Berliner and Biddle argued that 8th grade math comparisons have been unfair because, unlike students in other countries, most of OUR students do not take algebra in the 8th grade. Algebra items, however, make up only part of the international tests, and the results are virtually the same whether they are included or not. In the 1991 IAEP-2 math study, for example, the U.S. still would have scored BELOW the international average and trailed the leading countries by 16 to 18 percentage points (Lapointe, Mead, & Askew, 1992, pp. 39, 146).

Their response to me was baffling: "Somehow Stedman takes this simple demonstration of the effects of differences in curricula and opportunity-to-learn and converts it into a series of assertions that we did not make in TMC and do not believe."

As discussed, this was anything but a "simple demonstration" of curriculum differences; in fact, it was quite flawed. Furthermore, I have to ask: What "series of assertions"? I simply discussed Westbury's actual methods and findings that pertained to THEIR opportunity-to-learn claim and noted that they failed to discuss the 12th grade results which showed U.S. students at a serious mathematical disadvantage--even after curricular differences had been taken into account! As I discussed in my review of the international assessments (Stedman, 1994a), curriculum differences and opportunity-to-learn can only explain part of the U.S. international achievement deficiency. Furthermore, the lack of U.S. curriculum coverage, particularly in mathematics, often reflects our less demanding and weaker academic program, and so does not excuse our low achievement.

By the way, Berliner and Biddle also violated the Principle of Control in their public vs. private schools graph--p. 123--when they showed that public school students who take advanced math courses slightly outperformed private school students. This does NOT prove, however, as they asserted, that the public-private difference is simply a matter of curriculum--the public school advanced math takers are a select, elite group. Here again, they failed to disentangle curriculum and class effects. Furthermore, although their graph came from AFT research reported by Albert Shanker (1991, p. 10), they never mentioned that he concluded that both sectors were achieving poorly! (Although I agree with their general point that the private vs. public school achievement gap has been overblown, I wouldn't characterize the gaps as generally "small" as they did--in the 1990s NAEP comparisons they have often been substantial, but probably not that much more than would be expected given that private schools have a more upscale student body. I also think that Shanker's conclusion is an intriguing one that is well worth exploring further.)

## VALIDITY AND SAMPLING BIAS IN THE INTERNATIONAL ASSESSMENTS

Finally, they offered a series of arguments about the appropriateness and validity of the international assessments which are not supportable. In the first one, Berliner and Biddle are caught in a Catch-22. They argue that the international tests have not measured "the unique values and strengths of American education", including "creativity, initiative, and independence of thought in students"--yet at the same time, their book criticizes today's schools for lacking these very features. They are clearly concerned that neoconservative strategies, such as work intensification and national standards, are dominating schooling and propose numerous progressive alternatives (cooperative learning, project method, etc.) designed to rectify the situation and enhance creativity and initiative.

There is also a certain hubris in asserting that "American" education is "uniquely" focused on such things. As I noted, Japanese elementary students have rich curricular and extra- curricular activities--calligraphy, sewing, hands-on math and science activities, group problem-solving, electronics, dance, musical training, play, reading, physical exercise, cooperative learning, school jobs, etc. Without explanation, however, they labeled this as one of my "stranger" assertions! Furthermore, our breath of focus hardly excuses our low levels of achievement and knowledge--our schools, parents, and policy makers all clearly value high levels of achievement.

They also argued that sampling bias is a major problem for the international assessments, claiming that the assessments compare the broad mass of U.S. high school students to select samples in high-status high schools overseas (p. 54). Others have claimed similarly that our average student was compared to an elite, university-bound group of European students. This is an old criticism, however, emerging out of the first round of IEA international assessments in 1964 and 1970-71. Even then, the severity of the sampling problem varied by country and subject. In mathematics, the assessment deliberately sampled seniors who were taking math as part of a college-preparatory sequence. This narrowed the U.S. selection to college-bound students (only 18%) and thus avoided an unwarranted mass-to-elite comparison.

Their claim is even less applicable to the second international IEA math study, where many countries had 12th grade math enrollment rates similar to that of the U.S. (which was only 13%). Furthermore, most of these countries outperformed the U.S. by a considerable margin (Stedman, 1994a). Even some of the countries with higher enrollment rates matched or outperformed the United States. Hungary, for example, scored about the same as the small U.S. elite in several areas even though it enrolled half its students! In the second international science study in the mid-1980s, the U.S. actually had more selective 12th grade enrollments than most countries and still achieved more poorly in chemistry, physics, and biology. (Their example of a Japanese teacher's comments about sampling problems is a red herring. It has nothing to do with the major international assessments--IEA or ETS's IAEP.)

Critics have made too much of the variations in high school enrollments. Most of the assessments have involved 9- to 14-year- olds, ages when education is compulsory in developed countries and nearly 100% of the students are represented. Unfortunately, these are also the ages where the U.S. has struggled in several subjects.

On another point. I, too, am concerned about the newsmedia's inadequate coverage of the international assessments, but that does not prove that U.S. schools "stack up very well".

One of the worst features of Berliner and Biddle's response is that they repeatedly retreat from or even misrepresent their own position! As to variability, they now claim:

> Stedman asserts that we had argued that overall variability in achievement among students should be greater in our country, but we did not argue for such an effect.
>
> Yet, there's what they wrote in their book:
>
> Together these two problems [disparities in student wealth and inequities in funding] mean that scholastic achievements will vary far more in the United States than in other countries (p. 58).
>
> and
>
> To state this issue succinctly, the achievement of students from American schools is a LOT more variable than is students achievement from elsewhere (p. 58, emphasis original).

As I noted, the evidence does not bear out this sweeping contention. In fact, the 1991 IAEP math and science studies showed our variability was similar to that of other nations and less than that of Taiwan and Korea, the leading performers.

I have no trouble with the implication of the states-to-nation comparison they presented. Clearly, there are enormous regional variations in U.S. achievement and it is always useful to look at disaggregations of data for other patterns. What I was concerned about was their failure to inform the reader that this comparison had been labeled "experimental" and was technically problematic. (Contrary to their assertions in their response, they did not report in their book the details of the data or how the comparison was conducted!) Furthermore, when even our best state scores (those from a few typically high-scoring mid-Western states), are only at the AVERAGE level of Taiwan and Korea, we have cause for concern. Both aggregated and disaggregated scores indicate a serious problem in mathematics.

Finally, although minority and low-income students achieve relatively poorly, that remains insufficient to explain our generally low achievement. As I explained, the math deficit is not simply a minority student problem. In 1992, only 30% of WHITE U.S. 8th graders demonstrated NAEP math proficiency while over a quarter did not even make the basic level. Nor are our problems due to low-achievers. Even our top half have not kept pace internationally in math and science (Stedman, 1994a). Why do their "minds boggle" over such straight-forward explanations?

Instead of dealing with this evidence, they twisted my explanation into an argument that I claimed the low scores of minority students had no impact on average scores! Which is, of course, ridiculous. The point is that a major math problem and gap remains even when one looks at (disaggregates) other portions of the data--such as white students and the top half. It is also worth noting that, with the same demographics, U.S. reading scores are quite strong internationally.

Berliner and Biddle should have admitted that they selectively reviewed the international evidence, presenting only a couple of scattered pieces that supported their viewpoint. I invite readers to read my comprehensive analysis of the international assessments, in which I report the major findings and discuss the assessments' strengths and weaknesses (Stedman, 1994a).

## SWEEPING CLAIM #4: THE EDUCATIONAL CRISIS IS MANUFACTURED

> In addition, Stedman asserts that we made another "sweeping claim," that "the general education crisis is [merely] a right-wing fabrication," although he provides no citation to justify this charge. Again, this misrepresents what we wrote.

This is remarkable. This claim of theirs--that the general education crisis is not real and was manufactured by right-wing forces--is one of the central arguments of their entire book.

My review, however, was not focused on their political assertions but rather on their claim that the achievement crisis is a myth. Hence, my title "The Achievement Crisis is Real" and my extensive review of the achievement evidence in my section, "Low Achievement".

Let me be clear. I believe that right-wing forces have been attacking the public schools and EXPLOITING the evidence (and have been aided by a mix of social forces), but there is also extensive, credible evidence that there is a real achievement crisis, something Berliner and Biddle continue to deny. They have still never dealt directly with the actual evidence about low achievement.

Nevertheless, let us consider their charge. Note here that they had to add the word "merely" to my quote before discussing it. Does my statement really misrepresent what they wrote?

Let's quote and cite them from several places. First, begin with the title: THE MANUFACTURED CRISIS. Manufactured? By whom? Well, as they stated in their response "right-wing ideologues gained access to the White House with the election of Ronald Reagan, and in our book we detailed their influence on White House education policy." Here's how they explained the manufactured crisis and the lack of real evidence:

> We began our book by noting that throughout most of the Reagan and Bush years, the White House led an unprecedented and energetic attack on America's public schools, making extravagant and false claims about the supposed failures of those schools, and arguing that those claims were backed by "evidence." . . .
>
> No such White House attack on public education had ever before appeared in American history--indeed, even in the depths of the Nixon years the White House had not told such lies about our schools. Since the attack was well organized and was led by such powerful persons-- and since its charges were shortly to be echoed in other broadsides by leading industrialists and media pundits--its false claims have been accepted by many, many Americans. And these falsehoods have generated a host of poor policy decisions that have damaged the lives of hard-working educators and innocent students. In our book we labeled this attack "The Manufactured Crisis".

Ironically, they claimed that I was the one that was reducing complex realities to a "political slogan"!

In the introduction to their book, they point quite clearly to "organized malevolence" and "nasty lies," and alleged that "government officials and their allies were ignoring, suppressing, and distorting evidence" (xi). In their chapter 4, "Why Now?", they laid out their case that right-wing forces have manufactured the crisis, and titled various sections "The Entitlement of Reactionary Voices", "The Far Right", "The Religious Right", "The Neoconservatives" and "School-basing and Governmental Scapegoating". They argued that,

> Early in the 1970s, however, a number of wealthy people with sharply reactionary ideas began to work together to promote a right-wing agenda in America (p. 133).
>
> . . . these foundations have undertaken various activities to "sell" reactionary views: funding right- wing student newspapers, internships, and endowed chairs for right-wing spokespersons on American campuses. . . lobbying for reactionary programs and ideologues in the federal Congress (p. 133).

They were quite clear in arguing that the "Manufactured Crisis was not merely an accidental set of events or a product of impersonal social forces" (p. 9) but involved a "serious campaign by identifiable persons to sell Americans the false idea their public schools were failing and that because of this failure the nation was at peril."

They themselves, therefore, have made it quite clear that they believe that the achievement crisis was a right-wing fabrication.

## THE MANUFACTURED CRISIS REVISITED

In my review, I only touched the tip of the iceberg as far as their errors and distorted evidence went. One of the most egregious examples of misleading and selective presentation was their handling of opinion data on schools. It is worth exploring at length for it is both a crucial piece of evidence and argumentation and illustrates how they select confirmatory evidence and ignore disconfirmatory.

## PARENTAL (DIS)SATISFACTION WITH THE SCHOOLS

In a compelling comparison, Berliner and Biddle pointed out that opinion about the national status of education, which was supposedly influenced by the conservative assault, is negative, and then claimed that parents' judgments of their community's and children's schools, which were supposedly based on local information, are quite positive. Here we have an important piece of evidence that goes right to the heart of their argument about a manufactured crisis. Berliner and Biddle argued that the negative opinions about national conditions are "stereotypic" reflecting "rumors" and "bad portrayals" in the "popular press" and are, in essence, manufactured by right-wing neoconservative critics, whereas the positive opinions about local schools are based on "personal experience, direct observation, informed judgment, and discussions with others" (p. 112). In particular, parents of school-age children will have "first-hand, direct knowledge" and their opinions are "more likely to reflect reality." Thus, according to this argument, our schools are actually in good shape because that's what parents and local opinion says.

At one level, this is a very curious argument for them to be making given their interest in sweeping educational changes. If it were true, it spells disaster for their own reform agenda. It would mean that parents are quite satisfied with what is going in their local schools and there would be little justification for progressive reforms.

Before reviewing the actual data, let us consider a different perspective on why opinion about local schools might be more positive. Andrew Coulson (1994) makes an intriguing counter- argument--namely, that citizens are better informed about the national condition of education than they are about the local one. Every few years, for example, the National Assessment of Educational Progress reports on students' knowledge and skills in major academic areas--history, civics, geography, reading, mathematics, writing, etc.--and the findings are widely distributed in the media. It could well be that, if parents had the same kind of detailed achievement information about local students' knowledge and performance, they would be just as critical of their local schools.

I think Coulson is on to something. Few parents ever visit classrooms, particularly at the high school level, or shadow students throughout a day; few have ever actually observed what goes on inside the schools. Few districts routinely gather and report to the local media and community information about what students know and can do. In most communities, there is no systematic testing and reporting of high school students' knowledge in the key academic subjects. (I am referring here to curriculum-based exams in Algebra II, English Literature, U.S. History, Civics, Spanish 2, etc. and not generic, commercial standardized tests of reading, math, and social studies that are sometimes reported.)

If the results on such exams were regularly reported, and if parents routinely spent time in classrooms during the day, judgments of local schools could well be more negative. (Similarly, if parents were familiar with the many ethnographies of school conditions that were produced over the past decade, they might be decidedly more critical of their local schools.)

My primary concern here, though, is with the actual evidence and how Berliner and Biddle presented it. For over 25 years, Gallup and Phi Delta Kappa have surveyed the educational opinions of a national representative sample of adults, including public school parents. They have repeatedly asked respondents to rate the schools on the A, B, C, D, and Fail grading scale.

Berliner and Biddle used this data to claim that public school parents are "well satisfied with their schools" and "rate them highly" (p. 114). But, in presenting the data, they combined A and B ratings, which thus inflated the positive ratings, and omitted grades of C entirely! Their graph of parental opinion was an unusual one, therefore, in that it contrasted A/B ratings with D/F ratings and left out Cs entirely (p. 113). The result was a skewed comparison. (Their graph also contained a error-- what they labeled as the adult sample's opinion of local schools was actually that from respondents with no children in schools!)

Contrary to their selective approach, I here present tables of the 1993 results complete with each of the grades, A through Fail, so that readers can inspect them (Elam, Rose, & Gallup, 1993). The first table gives the ratings by all respondents, the second gives the ratings of public school parents.

**1993 RATINGS--ALL RESPONDENTS**

| | A | B | C | D | Fail | Don't know |
|---|---|---|---|---|---|---|
| Nation's Public Schools | 2 | 17 | 48 | 17 | 4 | 12 |
| Public Schools in this community | 10 | 37 | 31 | 11 | 4 | 7 |

**1993 RATINGS--PUBLIC SCHOOL PARENTS**

| | A | B | C | D | Fail | Don't know |
|---|---|---|---|---|---|---|
| Nation's Public Schools | 3 | 16 | 49 | 17 | 4 | 11 |
| Public Schools in this community | 12 | 44 | 28 | 12 | 4 | <.5 |
| School your oldest child attends (did not specify public or private) | 27 | 45 | 18 | 5 | 2 | 3 |

So what do we find? Public school parents certainly do rate local schools more highly than national ones--fewer Cs, Ds, and Fails, and more As and Bs. But look closely at the data. Only about a QUARTER of public school parents rate their oldest child's school an A, which is hardly a ringing endorsement. A quarter apparently have serious concerns about it, rating it C through Fail. (By 1995, this percentage had grown to over a third; see Elam & Rose, 1995). Furthermore, almost half the public school parents (44%) in 1993 expressed some displeasure with their community's schools, rating them C through Fail. (By 1995, this figure had grown to exactly half.)

Nonpublic school parents' responses were particularly revealing as the next table shows. They were quite critical of their local public schools. About 2/3 rated them C through Fail. Although one might argue that they are less familiar with the public schools, one could conversely argue that the reason they became private school parents is because they know all too well what local schools are like.

## 1993 RATINGS--NONPUBLIC SCHOOL PARENTS

| | A | B | C | D | Fail | Don't know |
|---|---|---|---|---|---|---|
| Nation's Public Schools | 6 | 9 | 48 | 15 | 12 | 10 |
| Public Schools in this community | 5 | 32 | 41 | 9 | 11 | 2 |

All this data hardly suggests that "American parents" are "well satisfied with their local schools" as Berliner and Biddle argued (p. 114).

Berliner and Biddle compounded the distortions by then claiming that

> What is amazing is that this high level of parental satisfaction with their local schools is growing and is actually HIGHER today than it was seven years ago (p. 112).

Although "satisfaction" (As & Bs) grew in the late 1980s, ratings in the 1990s leveled off. In fact, 1993 ratings were a point lower than those of 1991, and 1992 ratings were a point lower than those of 1986. By 1995, ratings had fallen back to 1986 levels (Elam & Rose, 1995).

In any event, how do they explain these trends? They don't bother to. A conservative critic might argue, however, that the reason satisfaction grew in the 1980s was because schools went back to the basics, raised standards, improved discipline, etc., but this interpretation is not considered by Berliner and Biddle. Interestingly, the increases in parental satisfaction took place in the aftermath of reforms generated by A NATION AT RISK. Was this a reflection of real improvement? Or of national activity and publicity influencing local opinions?

## WHO IS SITTING IN JUDGMENT?

Berliner and Biddle condemn several prominent educators for mistrusting positive parental opinion about their local schools. They wrote:

> Who are Doyle, Ravitch, Finn, and Stevenson to tell them they are wrong? (p. 114)
>
> In effect, these critics have proclaimed themselves part of an elite who, for the good of the nation, will be pleased to tell other Americans what they are to believe and how they are to act (p. 114).

But isn't that exactly what Berliner and Biddle have done in their 414 page book as they lay out a progressive reform agenda and critique the conservative approach, one that turns out to have much parental support?

Why do Berliner and Biddle only respect--and present-- parental opinion when it suits them and not respect it--or discuss it--in other areas? The PDK/Gallup opinion study that Berliner and Biddle relied on (Elam, Rose, & Gallup, 1993) reported that the overwhelming majority of respondents have favored, for a long time, national achievement goals and standards, requiring a standard exam to get a high school diploma, and using national tests to compare communities' achievement.

Other parental opinions also ran counter to their (and my!) preferred approach. In 1993, two-thirds of PUBLIC SCHOOL PARENTS favored English immersion for language minority students or even instruction at parents' expenses over bilingual education. Half supported longer school years. In 1995, three-fourths of public school parents favored a constitutional amendment to allow prayers to be spoken in public schools (Elam & Rose, 1995). These results were similar to those from 1984. Most preferred a moment of silence for silent prayer or contemplation rather than spoken prayer.

The 1995 poll also shows that parents continue to strongly support national exams and standards. Over 80% of public school parents support higher standards in the major academic subjects for promotion and for graduation. About 60% favor them even if it meant "significantly fewer students would graduate". About three-fourths even favor setting standards for kindergarten through 3rd grade. About two-thirds of public school parents favor using standardized, NATIONAL exams for promotion in THEIR OWN community schools.

Such parental opinions do not simply reflect the national conservative hegemony that emerged in the last decade during the Reagan and Bush administrations. Although support for such measures as national testing grew a bit in the 1980s, it has a long history (Elam, Rose, & Gallup, 1993). Way back in 1970, people were advocating NATIONAL tests to measure their community's achievement and, even in the mid-1970s, most were advocating that all students be required to pass a standard exam to receive a high school diploma--and this was well before the conservative onslaught occurred that Berliner and Biddle labeled the MANUFACTURED CRISIS. So the issue of parental opinion is a complex one.

This past year, the Phi Delta Kappa/Gallup poll explored the reasons parents rated their local schools higher than the nation's (Elam & Rose, 1995). Their answers were striking and challenge Berliner and Biddle's complacency about academic achievement. Given a list of 11 possible reasons, Elam and Rose reported that the parents made a "significant number-one choice: THE LOCAL SCHOOLS PLACE MORE EMPHASIS ON HIGH ACADEMIC ACHIEVEMENT" (p. 43, emphasis original). So, if Berliner and Biddle are right that local parents are in the know about their public schools, then they should also respect their opinions about emphasizing academic achievement.

One limitation of this finding, however, is that the parents generally agreed with each of the choices they were offered--with one notable exception, that their children's schools were better because they had more to spend per pupil. That exception has relevance for the next section.

## PROBLEMS IN LAKE WOEBEGONE

If parents truly were satisfied with their schools, it would undermine Berliner and

Biddle's case for reform. So they had to find some support in the data for their reform agenda. PDK/Gallup asked respondents an open-ended question: "What do you think are the biggest problems with which the public schools of this community must deal?" Here's how Berliner and Biddle characterized the findings:

> In fact, the biggest complaint that American parents indicated in the 1993 Gallup poll was that their local schools were not supported adequately. This complaint took precedence over their concerns about drug abuse, lack of discipline, fighting, violence, gangs, and a host of other real and imagined problems (p. 114).

This neatly fits the basic argument Berliner and Biddle are advancing, but is truly misleading. THERE WAS NO CONSENSUS IN PARENTAL OPINIONS ABOUT SCHOOL PROBLEMS. A lack of proper financial support was the most often mentioned problem, but ONLY 24% of the public school parents cited that. THE VAST MAJORITY CITED OTHER PROBLEMS. It is unclear that funding took "precedence" over other problems. Respondents were not asked to rank problems. Almost half of them (43%) were concerned about issues of order and behavior--15% cited discipline, 14% drugs, and 14% fighting, violence, and gangs.

I find it curious that they would label some of the problems "imagined". Why were they suddenly discounting certain parental opinions, given that it is supposedly informed opinion? (Interestingly, 10% of the public school parents reported they had no idea what the biggest problems were.) They didn't mention that those without children in school responded similarly to public school parents, which further undermines their argument about locally-informed opinions.

It is likely that 1991-1993 concerns over finances were partly influenced by national happenings--the 1992 Bush-Clinton election campaign that focused in part of support for education and Jonathan Kozol's book SAVAGE INEQUALITIES--rather than simply the "reality" of the local situation. The survey itself may also have played a part in inducing financial concerns in that there was a series of questions about educational expenditures--equal funding, the impact of money, support for poor communities, etc. (One hopes that those questions came after the question about biggest problems.)

By 1995, the mention of financial support had dropped in half to only 12%. A lack of discipline was mentioned just about as often (11%). Had local conditions changed so dramatically? Had schools suddenly received adequate funding? Or, had the national debate shifted?

Berliner and Biddle identified the opinions about problems as those of "parents"--but it was actually parents with children currently in the public schools. Parents of nonpublic students made different, and quite intriguing, comments about their community's public schools. A lack of proper financial support was NOT the problem they most often mentioned in 1993 (or 1995). Instead, they were most often concerned about a lack of discipline in the local schools (19%), the standards and quality of education (18%), and fighting, violence, and gangs (17%). Although Berliner and Biddle ignored them, their opinions about local schools are worth listening to as they were the ones who decided to remove their children from those schools--or not put them there in the first place.

Opinions about public schools and reform, I believe, reflect a complex, highly tangled interaction of parental experience with local schools, the spirited national debate over educational reform, and a growing conservative hegemony.

Instead of recognizing these complex influences on parental opinion, instead of respecting the opinions of all parents, it was far simpler for them to set up false dichotomies--parents vs. nonparents, national illusions vs. local realities, and manufactured crisis vs. high satisfaction.

In the end, the "problem" became those without children. Berliner and Biddle commented about public school parents:

> The major problem they face is trying to persuade those who do not have children in the schools to agree to pay their share of school taxes (p. 114).

Such a sweeping comment flies in the face of the very survey they were reporting on

(Elam, Rose &, Gallup, 1993). Two-thirds of the respondents WITHOUT children in school said they would be willing pay more taxes to improve the quality of public schools in poorer states and communities. That figure closely matches the 71% of public school parents who said they'd be willing. 59% of those without children in school said they'd be willing to pay more federal taxes to improve inner-city schools, just about the same as the 62% of public school parents. Those without children in school also gave similar responses as to the local schools' biggest problems--although a lack of proper financial support was first, drugs, discipline, and violence together garnered the lion's share of the concerns.

Berliner and Biddle then concluded their discussion of parental opinion with:

> Perhaps it is time for citizens without children to join parents and go into the schools to see for themselves what is actually happening there (p. 114).

Perhaps it is time for both groups (along with educational researchers) to do just that!

The main point I am making in this section is that opinions about local schools are nowhere near as strong as Berliner and Biddle argue--one can hardly describe it as "remarkable degree of consumer satisfaction" (p. 113) when half the public school parents are rating their community's schools C through Fail. What it suggests to me is that there is a deep well of dissatisfaction that could be enlisted in a movement toward progressive reform. But we must understand and respect the fact that public school parents have many conservative ideas about schooling and reform, shaped by national forces (and conservative propaganda) but grounded as well in local experiences.

## PROGRESSIVE REFORMS AND THE RIGHT-WING AGENDA

There should be little question that I basically agree with Berliner and Biddle's reform mission. As I wrote in the WASHINGTON POST review,

> Berliner and Biddle offer a welcome critique of the neoconservative agenda--privatization, national testing, gifted programs, and work intensification. They forcefully document the social problems plaguing our schools--from economic stagnation to poverty--and provide a useful compendium of alternative reform strategies--small schools, authentic assessment, equitable funding, and community involvement.

As a progressive educator, therefore, I'm sympathetic to their concerns. The ascendancy of the political right is troubling and could harm public education greatly. We do need to overhaul school financing systems and do more for low-income rural and urban students. We do need to critically examine neoconservative reform strategies and aggressively promote progressive alternatives.

Ultimately, though, the book suffers from being one-sided. While right-wing "organized malevolence" and government suppression of evidence make for good reading, they do not mean the educational crisis is a myth.

Berliner and Biddle were so intent, for example, on branding the major 1980s reform reports as ideologically conservative, that they even tarred thoughtful critiques of the schools by progressive educators. Their list of reports, for example, that were supposedly products of conservative ideologies and Human Capital theories included A PLACE CALLED SCHOOL by John Goodlad and HORACE'S COMPROMISE by Ted Sizer (p. 140).

## THE SUPPRESSION OF THE SANDIA REPORT

They were more on target when they described how the conservative political agenda shaped the Department of Education's WHAT WORKS? reports and how self-interested budget considerations may have led NSF to stand by a flawed study predicting a national shortfall of scientists (pp. 162-164). But then they went further and, without evidence, suggested that NSF stood its ground because the Reagan administration was interested in helping industrialists (p. 165).

In a more dramatic tale, they also alleged the Bush administration suppressed a major study of education-- the Sandia Report--because it contradicted official claims about the poor state of education, and would have set the achievement record straight (pp. 165-168). This story is an important one because the report formed the basis of several well-known articles challenging the notion of an educational crisis (see, e.g., Bracey, 1991; 1992) and Berliner and Biddle extolled its virtues (pp. 26, 354).

The report was rife with errors, however, which helped delay its publication and they overlooked its substantial shortcomings-- sloppy analysis of the SAT and international data and omission of key achievement data (Stedman, 1994b).

The allegation of suppression is a serious one and potentially libelous. Berliner and Biddle had an obligation to furnish the evidence for such charges IN their book and, in the interest of fairness, present alternative interpretations of the events-- particularly giving the viewpoints of those charged with suppressing. This they did not do. They simply alleged that administration officials subjected the report to "unprecedented" NCES and NSF reviews, yet it seems that the reports' authors were involved in requesting the reviews. In 1993, one of the authors, Robert Huelskamp, wrote that, "As our work unfolded in the spring of 1991, WE SUBJECTED a draft to peer review with the U.S. Department of Education, the National Science Foundation, and other researchers (most notably Gerald Bracey)" (Huelskamp, 1993, p. 719, emphasis added).

It has struck many observers as reasonable that a report on education created by Department of Energy analysts--not by educators--should be reviewed by education researchers at the National Center for Education Statistics, people who would be more conversant with the data. Berliner and Biddle offered no evidence that such a review was unprecedented (nor did the source they relied on--Tanner, 1993); indeed a major Energy report on the general condition of K-12 public schooling was itself something unprecedented. As one of its authors noted, it was a departure from previous efforts that had focused on analyses of postsecondary education and the training of scientists and mathematicians (Huelskamp, 1993, pp. 718-719).

Berliner and Biddle also wrote that "the report itself eventually appeared in the JOURNAL OF EDUCATIONAL RESEARCH--without fanfare, without even a listing of its authors!" (p. 159). In fact, Huelskamp (1993) first published a version of the report in PHI DELTA KAPPAN, one of the largest circulating educational journals, and informed readers that the "full report will be published in the May/June issue of the JOURNAL OF EDUCATIONAL RESEARCH" (p. 719). Furthermore, the entire issue of JER was devoted to the report and its front cover listed the authors' names--C.C. Carson, R. M. Huelskamp, and T.D. Woddall--in bold print!

Even though it took time for the final report to be released, its ideas were widely circulated much earlier. The authors themselves distributed drafts of the report even before the summer 1991 NCES and NSF reviews were completed (Miller, 1991, p. 32). Gerald Bracey (1991) used them as the basis of his first annual report on the condition of education that appeared in PHI DELTA KAPPAN back in 1991, an article that received widespread publicity, and he later credited them with helping change conservative critics' views of the achievement decline (Bracey, 1992). The report's authors also testified to Congress in the summer of 1991 and the printed testimony, including a synopsis of the report, was readily available (HEARINGS ON THE STATE OF EDUCATION, 1991).

To be sure, the entire episode is quite controversial. Miller (1991) reported that unnamed sources contended the authors were worried about possible reprisals (funding cut-offs), a GAO audit was conducted, several politically-charged statements were revised out of the draft, etc. Several sources did charge that the report was being buried because it conflicted with Bush administration educational policy and that the Congressional testimony was needed to get the message out. Administration officials countered that the report was delayed because it was undergoing an expert review process.

Whether it was suppressed, buried, delayed, or legitimately subjected to additional reviews (or several of the above!), such actions do not mean that the report's findings were valid and should be accepted. Berliner and Biddle claimed that NCES and NSF reviewers "dutifully detected trivial 'flaws' " (p. 167) ), but like Tanner (1993), they did not present the reviewers' findings or what was concluded about the nature and extent of the flaws. In fact, the reviewers raised serious, fundamental questions about the quality of the report, its data handling, and its conclusions.

(Tanner argued that the reviewers were opinionated and provided one example where some reviewer had unprofessionally written "Nuts" next to a passage on a Sandia draft (p. 292)-but a blunt opinion hardly invalidates what many reviewers found or what the summaries of the reviews concluded).

In his summary of NCES's review, Emerson Elliott (1991), the commissioner of NCES, described the problems as follows:

> The report appears to be highly selective in the information it presents. Information that is widely known and understood is not presented, and the data shown are consistently supportive of a picture of U.S. education in a positive light. This could give rise to criticisms that the report is a biased presentation instead of the "balanced" presentation that has been claimed.
>
> . . . the trends in educational performance among U.S. students are complex and not well-represented in this analysis. The oversimplification leads to simplistic interpretations.
>
> In many places in the report the findings and interpretations are not supported by the data presented.
>
> . . . the results of the science examinations in the NAEP are provided. The assertion is made that the trends shown are consistent with the results of exams in other subject matter areas. This is not the case, as demonstrated in numerous analyses of NAEP and other achievement data.
>
> The discussion of international comparisons on test scores reflects this problem as well. Many other international comparisons have been made, and some of the issues identified in the issue discussion on p. 94 have been addressed in studies. These findings should have been included for a more balanced discussion of U.S. student performance.
>
> A longitudinal component over the course of a year permitted comparison of what students were actually taught during a year and how they performed on those test items. The U.S. performance, unfortunately, was rather dismal.
>
> He concluded that the report contains:
>
> assertions that contradict what we know well from broadly grounded research conducted over a number of years with repeated replications using different databases
>
> misinterpretations of the data presented
>
> inappropriate policy conclusions [and]
>
> conclusions not well founded in the information presented.

The NSF review determined that "the report rests on a partial and flawed analysis" and that its conclusions are "not adequately supported" (House, 1991). The NSF reviewers (several not just one as Berliner and Biddle suggested) found "several major flaws" typified by a "lack of understanding of the data series used" and "unresolved conflicting interpretations" (House, 1991). They noted there were "dozens of flaws" and gave many examples, including the Sandia analysts' sweeping claim there wasn't ANY NAEP test that showed declines and their failure to recognize students' low achievement levels on the tests.

My own review concluded that the report was

> generally right about steady trends, but that it is seriously flawed by errors in analysis, insufficient evidence, mischaracterizations of the international data, and a failure to consider the evidence that U.S. students are performing at low levels. In spite of its findings, fundamental school reform is still warranted (Stedman, 1994b).

Interested readers can find a detailed treatment of the report's strengths and limitations in Stedman (1994b).

## SHAPING A PROGRESSIVE REFORM AGENDA

Berliner and Biddle also characterized the present national agenda as right-wing and neoconservative, but it was developed across the political and educational spectrum--by governors of both parties, teacher union leaders, and state school superintendents. While right-leaning, it contains a complex mixture of reforms. Even the national Goals 2000 program includes such long-time progressive objectives as parental participation and ensuring children come to school ready to learn.

Let me be clear. I have no doubt that right-wing forces have organized an assault on the public schools; that conservative school critics exploited the evidence and exaggerated the decline. I was, for example, an early critic of the NATION AT RISK for misusing data, exaggerating the decline, and ignoring equity issues (Stedman & Smith, 1983; see also Stedman & Kaestle, 1985). But just as conservative critics were wrong to argue that we were in a massive decline and needed to return to traditional schooling, so too, progressives such as Berliner and Biddle are now wrong to suggest that our schools are achieving well and that concerns about students' levels of knowledge are unfounded.

As I explained in my WASHINGTON POST review (Stedman, 1995), progressives should be willing to admit that achievement is low. But that does not mean embracing a conservative agenda or calling for the U.S. to be #1 in the world in math and science, as the nation's Goals 2000 program does. Nor does it mean calling for the schools to go back to old-fashioned, regimented teaching. The existing curriculum is already too facts-based and memory- driven and is not working. As I wrote in the POST review:

> An historical perspective helps here. Conservatives often blame the decline of excellence on 1960s liberalism, but students' achievement and general knowledge were low even in the 1940s and 1950s--a clear indication traditional practices have never been very successful. Such persistent failure strengthens the case for a sweeping, progressive restructuring of schools.

Berliner and Biddle, therefore, missed a great opportunity to strengthen their own case for progressive reform. By combining the progressives' call for cooperative learning and rich curricula along with the conservatives' emphasis on high levels of knowledge, we would be far more likely to develop reflective, well-informed students. (Note as well that thoughtful conservatives are also calling for innovative teaching methods, an engaging, challenging curriculum, and an end to tracking.) A far more compelling case for reform could be made--and one that could garner more universal support--when we explain that traditional methods have failed and that even children of the middle-class are often not mastering important academic knowledge.

I invite readers to compare my analyses of the condition of educational achievement with theirs (see bibliography). Judge for yourselves who has produced the balanced, careful treatment of the data; who is willing to acknowledge the complexity of the data and achievement patterns, and who is working hard at understanding the evidence rather than trying to fit it into one neat, pat story. Although we should be concerned about the growing influence of right-wing politics, let us also respect the evidence; the achievement crisis remains real and the need for fundamental school reform remains great.

## References

Anderson, L., Jenkins, L., Leming, J., MacDonald, W., Mullis, I., Turner, M., & Wooster, J. (1990). THE CIVICS REPORT CARD. Princeton, N.J.: Educational Testing Service.

Applebee, A. N., Langer, J. A., Mullis, I. V. S., & Jenkins, L. B. (1990). THE WRITING REPORT CARD, 1984-1988. Princeton, N.J.: NAEP.

Applebee, A., Langer, J., Mullis, I., Latham, A., & Gentile, C. (1994). NAEP 1992 WRITING REPORT CARD. Washington, D.C.: National Center for Education Statistics.

Berliner, D., & Biddle, B. (1995). THE MANUFACTURED CRISIS: MYTHS, FRAUD, AND THE ATTACK ON AMERICA'S PUBLIC SCHOOLS. New York: Addison-Wesley.

Berliner, D., & Biddle, B. (1996). Making molehills out of molehills: Reply to Lawrence Stedman's review of THE MANUFACTURED CRISIS. EDUCATION POLICY ANALYSIS ARCHIVES, 4(3). http://seamonkey.ed.asu.edu/epaa/

Bracey, G. (1991). Why can't they be like we were? PHI DELTA KAPPAN (October), 105-117.

Bracey, G. (1992). The second Bracey report on the condition of public education. PHI DELTA KAPPAN (October), 104-117.

Carpenter, T. P., Lindquist, M. M., Brown, C. A., Kouba, V. L., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP assessment of mathematics. ARITHMETIC TEACHER (December), 38-41.

Carson, C. C., Huelskamp, R. M., & Woodall, R. D. (1993). Perspectives on Education in America. THE JOURNAL OF EDUCATIONAL RESEARCH, 86 (May/June), 259-310.

Coulson, A. J. (1994). A Response to John Covaleskie. EDUCATION POLICY ANALYSIS ARCHIVES, 2(12). http://seamonkey.ed.asu.edu/epaa/

Elam, S., & Rose, L. (1993). The 27th annual Phi Delta Kappa/Gallup Poll of the Public's attitudes toward the public schools. PHI DELTA KAPPAN, 77(1), 41-56.

Elam, S., Rose, L., & Gallup, A. (1993). The 25th annual Phi Delta Kappa/Gallup Poll of the Public's attitudes toward the public schools. PHI DELTA KAPPAN, 75(2), 137-152.

Elliott, E. (1991). Review of the Sandia National Laboratory report on education. Letter to Richard E. Stephens, Associate Director for University and Science Education, Department of Energy from the Acting Commissioner, National Center for Education Statistics. Washington, D.C.: U.S. Department of Education.

Gilovich, T. (1991). HOW WE KNOW WHAT ISN'T SO: THE FALLIBILITY OF HUMAN REASON IN EVERYDAY LIFE. New York: The Free Press.

HEARINGS ON THE STATE OF EDUCATION. (1991). Hearings before the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor, House of Representatives, One Hundred Second Congress. Hearings in Washington, D.C., May 1 and 3, and July 18, 1991. Serial No. 102-28. Washington, D.C.: U.S. Government Printing Office. ISBN 0-16-035543-5.

House, P. (1991). Review of the Sandia National Laboratory report on education. Letter to Richard E. Stephens, Associate Director for University and Science Education, Department of Energy, from the Director of the Division of Policy Research and Analysis, NSF. Washington, D.C.: National Science Foundation.

Huelskamp, R. (1993). Perspectives on education in America. PHI DELTA KAPPAN, 74(9), 718-722.

Kozol, J. (1991). SAVAGE INEQUALITIES: CHILDREN IN AMERICA'S SCHOOLS. New York: Crown Publishing.

Lapointe, A., Mead, N., & Askew, J. (1992). LEARNING MATHEMATICS. Princeton, New Jersey: Educational Testing Service.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district results to national norms: The validity of claims that "everyone is above average". EDUCATIONAL MEASUREMENT: ISSUES AND PRACTICE (Fall), 5-14.

Miller, J. (1991, October 9). Report questioning 'crisis' in education triggers an uproar. EDUCATION WEEK, p. 1, 32.

Mullis, I. V. S., Dossey, J. A., Foertsch, M. A., Jones, L. R., & Gentile, C. A. (1991a). TRENDS IN ACADEMIC PROGRESS. Washington, D.C.: U.S. Government Printing Office. (ED 338 720)

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1991b). THE STATE OF MATHEMATICS ACHIEVEMENT. Washington, D.C.: U.S. Department of Education.

Olson, L. (1995, February 8). Students' best writing needs work, study shows. EDUCATION WEEK, 5.

Shanker, A. (1991, Fall). Do private schools outperform public schools? AMERICAN EDUCATOR, 8-15, 40-41.

Stedman, L. C. (1993). The condition of education: Why school reformers are on the right track. PHI DELTA KAPPAN, 75 (3), 215-225.

Stedman, L. C. (1994a). Incomplete explanations: The case of U.S. performance in the international assessments of education. EDUCATIONAL RESEARCHER, 23(7), 24-32.

Stedman, L. C. (1994b). The Sandia Report and U.S. achievement: An assessment. JOURNAL OF EDUCATIONAL RESEARCH (January- February), 133-146.

Stedman, L. C. (1995, November 5). Putting the system to the test. [Review of THE MANUFACTURED CRISIS.] Education Review section of the Washington Post, 16-17.

Stedman, L. C. (1996). The achievement crisis is real. [Review of THE MANUFACTURED CRISIS.] EDUCATION POLICY ANALYSIS ARCHIVES 4(1). http://seamonkey.ed.asu.edu/epaa/

Stedman, L. C., & Kaestle, C. F. (1985). The test score decline is over: Now what? PHI DELTA KAPPAN, 67(3), 204-210.

Stedman, L. C., & Kaestle, C. F. (1991a). The great test score decline: A closer look. In C. F. Kaestle, H. Damon-Moore, L. C. Stedman, K. Tinsley, & W. V. Trollinger (Eds.), LITERACY IN THE UNITED STATES (Chapter 4). New Haven: Yale University Press.

Stedman, L. C., & Kaestle, C. F. (1991b). Literacy and Reading Performance in the United States from 1880 to the present. In C. F. Kaestle, H. Damon-Moore, L. C. Stedman, K. Tinsley, & W. V. Trollinger (Eds.), Literacy in the United States (Chapter 3). New Haven: Yale University Press.

Stedman, L. C., & Smith, M. S. (1983). Recent reform proposals for American education. CONTEMPORARY EDUCATION REVIEW, 2(2), 85-104.

Tanner, D. (1993). A nation 'truly' at risk. PHI DELTA KAPPAN, 75(4), 288-297.

Viadero, D. (1995, September 13). Book that bucks negative view of schools stirs debate. EDUCATION WEEK, p. 8.

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? EDUCATIONAL RESEARCHER, 21(5), 18-24.

Williams, P., Lazer, S., Reese, C., & Carr, P. (1995). NAEP 1994 HISTORY: A FIRST LOOK. Washington, D.C.: U.S. Department of Education.

Williams, P., Reese, C., Campbell, J., Mazzeo, J., & Phillips, G. (1995). NAEP 1994 READING: A FIRST LOOK. Washington, D.C.: U.S. Department of Education.

Williams, P., Reese, C., Lazer, S, & Shakrani, S. (1995). NAEP 1994 GEOGRAPHY: A FIRST LOOK. Washington, D.C.: U.S. Department of Education.

Wirtz, W. et al. (1977). ON FURTHER EXAMINATION. New York: College Board.

## About the Author

**Lawrence C. Stedman**

stedman@binghamton.edu

Lawrence C. Stedman is Associate Professor of Education at the State University of New York at Binghamton. His Ph.D. is from the University of Wisconsin at Madison in Educational Policy Studies with a minor in Sociology. He has worked as a school district policy analyst, secondary school teacher, VISTA volunteer, and educational researcher. He has a keen interest in equal opportunity and school reform. His dissertation and early articles centered on effective schools research and the reform reports of the early 1980s. He has helped evaluate ESL, minority achievement, merit pay, and dropout intervention programs.

More recently, his research has focused on the general condition of education and its implications for policy-making. He has written articles on the test score decline, literacy trends, the international assessments, and the Sandia Report. He is currently investigating historical trends in students' and adults' general knowledge. It is the outgrowth of a book he helped author with Carl Kaestle and others on the history of the U.S. reading public (*Literacy in the United States: Readers and Reading Since 1880*, Yale University Press, 1991). This new research has been funded by a SUNY Faculty Research Grant and Fellowship and by a National Academy of Education Spencer Foundation post-doctoal fellowship.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 8** **April 21, 1996** **ISSN 1068- 2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# Developmentalism: An Obscure but Pervasive Restriction on Educational Improvement

**J. E. Stone**

**East Tennessee State University**

*STONEJ@EDUSERV.EAST-TENN-ST.EDU*

## Abstract

Despite continuing criticism of public education, experimentally demonstrated and field tested teaching methods have been ignored, rejected, and abandoned. Instead of a stable consensus regarding best teaching practices, there seems only an unending succession of innovations. A longstanding educational doctrine appears to underlie this anomalous state of affairs. Termed developmentalism, it presumes "natural" ontogenesis to be optimal and it requires experimentally demonstrated teaching practices to overcome a presumption that they interfere with an optimal developmental trajectory. It also discourages teachers and parents from asserting themselves with children. Instead of effective interventions, it seeks the preservation of a postulated natural perfection. Developmentalism's rich history is expressed in a literature extending over 400 years. Its notable exponents include Jean Jacques Rousseau, John Dewey, and Jean Piaget; and its most recent expressions include "developmentally appropriate practice" and "constructivism." In the years during which it gained ascendance, developmentalism served as a basis for rejecting harsh and inhumane teaching methods. Today it impedes efforts to hold schools accountable for student academic achievement.

Over the past thirteen years American public schools have been subjected to an increasing barrage of criticism. The chief object of complaint has been their continuing failure to equip students with the academic and workplace skills needed in an era of increasing economic competition.

Recent expressions evidence a growing public impatience. In an April 1993 statement, U. S. Secretary of Education Richard Riley commented: "A watered down curriculum and low expectations for too many of our students prevent them from meeting high standards" (Riley, 1993). A September 1993 report by the National Center for Education Statistics found that 16 to 20 percent of the U. S. adults who perform at the lowest levels of reading, writing, and arithmetic were high school graduates (Kirsch, Jungblut, Jenkins & Kolstad, 1993). In November of 1993, the U. S. Department of Education reported that in comparison to their peers in other industrialized countries, gifted American students rank near the bottom in math and science achievement (Kantrowitz & Wingert, 1993). In September of 1994, the American Legislative Exchange Council (ALEC, 1994) disclosed that since the Nation at Risk report in 1983 there has been little change in the achievement levels of public school students despite a 43% increase in real dollar expenditures. Near the end of 1994, the Organization for Economic Co-operation and Development (OECD, 1994) described the quality of American education as a major threat to the future economic well- being, productivity, and competitiveness of the U. S. In April of 1995, Business Week (Mandel, Melcher, Yang & McNamee, 1995) declared that businesses find too many job applicants unable to read, write, or do simple arithmetic and that Americans are "fed up" with their public schools.

Berliner and Biddle (1995) and various other commentators (Bracy, 1996; Westbury, 1992) have attempted to defend the public schools' record by offering a more sympathetic interpretation of the available evidence. However, a recent review of Berliner and Biddle (Stedman, 1996a) and the ensuing exchange between Berliner, Biddle and Stedman (Berliner & Biddle, 1996; Stedman, 1996b) demonstrates that reinterpretation of school and student performance data is unlikely to convince knowledgeable observers that the ongoing criticisms of public schooling are "manufactured" or otherwise off target.

Despite these mounting concerns, schools have largely ignored the availability of a number of teaching methodologies that seem capable of producing the kind of achievement outcomes demanded by the public. They are experimentally validated, field tested, and known to produce significant improvements in learning. Instead, the schools have continued to employ a wide variety of untested and unproven practices which are said to be "innovative" (Carnine, 1995; Marshall, 1993). In particular, teaching practices such as mastery learning and Personalized System of Instruction (Bloom, 1976; Guskey & Pigott, 1988; Kulik, Kulik & Bangert-Drowns, 1990), direct instruction (Becker & Carnine, 1980; White, 1987), positive reinforcement (Lysakowski & Walberg; 1980, 1981), cues and feedback (Lysakowski & Walberg, 1982), and the variety of similar practices called "explicit teaching" (Rosenshine, 1986), are largely ignored despite reviews and meta- analyses strongly supportive of their effectiveness (Ellson, 1986; Walberg, 1990, 1992). Yet methodologies such as whole language instruction (Stahl & Miller, 1989), the open classroom (Giacomia & Hedges, 1982; Hetzel, Rasher, Butcher, & Walberg, 1980; Madamba, 1981; & Peterson, 1980), inquiry learning (El-Nemr, 1980), and a variety practices purporting to accommodate teaching to student diversity (Boykin, 1986; Dunn, Beaudrey, & Klavas, 1989; Shipman & Shipman, 1985; Thompson, Entwisle, Alexander, & Sundius, 1992) continue to be employed despite weak or unfavorable findings or simply a lack of empirical trials.

Equally surprising is the observation that many of the ignored and rejected methodologies are quite similar to those that have been found effective and are routinely used by special educators and school psychologists (Hallahan, Kauffman, & Lloyd, 1985; Hammill & Bartel, 1990; Wang, Reynolds & Walberg, 1987). In many instances, the otherwise unused practices are successfully implemented but only after a student has been identified as disabled.

### Methods Texts and Experimental Research

A sampling of popular textbooks used in regular education teaching methods courses offers what may be a reason for this anomalous state of affairs. Widely used textbooks--in the present report, elementary, middle, and secondary teaching methods texts that have been revised repeatedly, some over thirty and forty years (Armstrong & Savage, 1994; Callahan, Clark, & Kellough, 1992; Clark & Starr, 1991; Henson, 1993; Jacobsen, Eggen, & Kauchak, 1993; Kim & Kellough, 1995; Lemlech, 1994; Ornstein, 1992; Sheperd & Ragan, 1992)--give little weight to experimentally demonstrated results as a basis for identifying effective teaching practices. Instead, they present an eclectic assortment of approaches colored by distinct distaste for methods that are structured, teacher- directed, and result- oriented--characteristics that exemplify the experimentally vindicated approaches to teaching. Lemlech's (1994) account is typical:

> In classrooms where students are given little opportunity to choose what they will learn, how they will learn, and the way

> in which they will be evaluated for learning, there is a greater likelihood that the classroom is structured through intrinsic rewards, incentive programs, and normative evaluation. As a consequence, learning will become joyless. There is also a tendency in these classrooms to overemphasize repetition, drill, and commercially produced dittos for practice materials. Some believe this to be prevalent in low socio-economic and low achieving classrooms, and as a consequence it may the cause of negative motivation patterns. (p. 91)

Instead of empirically grounded recommendations as to best practices, the methods texts suggest a personalized and intuitive approach to instruction built around teacher experience,circumstances, and sensitivity to student needs. Ornstein's (1992) advice exemplifies this view:

> In considering what is best for you, you must consider your teaching style, your student's needs and abilities, and your school policies. Asyou narrow your choices, remember that approaches overlap and are not mutually exclusive. Also remember that more than one approach may work for you. You may borrow ideas from various approaches and construct your own hybrid. The approach you finally arrive at should make sense to you on an intuitive basis. Don' let someone impose his or her teaching style or disciplinary approach on you. Remember, what works for one person (in the same school, even with the same students) may not work for another person. (p. 129)

In essence, these methods texts acknowledge research as a foundation for educational practice but give it little weight in formulating a conclusion about the practices most likely to produce results. Neither do they encourage the reader to rely on research as a basis judging the quality of teaching practices. They seem to wear the mantle of science but oddly neglect its substance and purpose.

The same emphasis on teaching shaped by innovation and sensitivity to student differences is quite evident in the catalogues of publishers that target teachers and teacher educators. The titles and descriptions of offerings by Heinemann (1995) and National Education Association (1995), for example, both reflect a market preference for the new and innovative and a market indifference to the empirically grounded or to the tried and true.

The varied and ever-mutating body of scholarship referenced by the textbooks implies the kind of ongoing refinement and revitalization characteristic of scientifically informed practice. Yet their recommendations with respect to teaching do not reflect the kind of consensus that would be expected to emerge as recent advancements are built onto established findings (Stanovich, 1992, 1993). Empirical findings are at best an imperfect guide to practice; but as they cumulate and converge, they do yield important clues. At the least, they reveal that certain findings tend to repeat themselves. The impression conveyed by the present textbooks, however, is that learning's relationship to teaching is largely idiosyncratic and unpredictable. That which is true for one teacher, teaching one lesson, to one set of students is not a valid guide for others.

Neither do these textbooks acknowledge the unique value of experimental trials. The distinctive value of experimental evidence is understood throughout the scientific community (Cook & Campbell, 1979), and experimentation as a guide to effective teaching practice has been recognized by the educational community for more than thirty years (Campbell & Stanley, 1963). Yet the methods texts are silent on the matter. Here again although the fallibility of empirical evidence must be acknowledged, it must also be said that the well conceived experiment offers more convincing evidence of whether a teaching method works than a report offering only description or correlation. Dismissing experimental findings on the grounds that offer only good but not certain evidence of pedagogical effectiveness is to fallaciously make the perfect the enemy of the good.

Given the market success of these textbooks and the teaching profession's apparent comfort with such an orientation, it is not difficult to see how schools continue to respond to the public call for better results with untested innovations (Carnine, 1995). Seemingly the education community has neither a scientifically founded consensus about best practices nor a recognition that experimental evidence would be integral to the formation of such a consensus. In the absence of attention to experimental trials, teaching innovations lacking demonstrated effectiveness can come into vogue on the strength of publicity and marketing only to later be bypassed by more of the same (Armstrong, 1980; Carnine, 1993; Marshall, 1993). In truth, continual innovation may have become a way of coping with public criticism. New practices are incongruously piled onto the old as consultants, school boards, superintendents, and teachers come and go (Armstrong, 1980). Criticisms that are behind the curve can be ignored because they are no longer relevant. Criticisms of the latest innovations can be ignored because they are premature and intolerant of innovation.

## The Influence of Developmentalism

The thesis advanced in the following is that a longstanding but poorly recognized educational doctrine underpins the neglect of experimental evidence found inmethods textbooks and in the attempt to find more effective teaching methods. It is a doctrine that pervades teacher education and one that disposes the teaching profession to favor certain practices and to ignore others regardless of empirically demonstrated merit. Termed "developmentalism" (Stone, 1991, 1993a, 1994), it is a form of romantic naturalism that inspires teacher discomfort with any practice that is deemed incompatible with natural developmental processes (Binder & Watkins, 1989). It is a view that acquired popularity as a grounds for rejecting the often harsh formalist teaching methods of the eighteenth and nineteenth centuries (Ravitch, 1983; Riegel, 1972). Today it poses an obscure but powerful restriction on scientifically informed educational improvement and more broadly on teacher and parent efforts to influence the developing child.

Developmentalism's clearest present-day expressions include the "child centered" or "progressive" teaching seen in Canadian schools (Freedman, 1993), the "progressivism" or "Plowdenism" seen in the British Primary Schools (Alexander, Rose, & Woodhead, 1992), and the "developmentally appropriate practice" advocated by early childhood educators (Carta, Schwartz, Atwater & McConnell,1991). The learner-centered teacher education favored by National Education Association is another expression, one that is widely known and well regarded in colleges of education (Darling-Hammond, Griffin & Wise, 1992).

Discovery learning is predicated on developmentalism (Bruner, 1966) and so is the increasingly popular constructivism (Brooks & Brooks, 1993). Although constructivism employs a distinctive terminology and a more credible theoretical foundation, its major precepts are largely those advanced by John Dewey (1916/1963) at the turn of the century and discredited in the nineteen fifties. Dewey's "progressive education" (Dewey, 1938/1963) is the best known historic form of developmentalism and one whose present day influence is remarkably underestimated. "Reflective thinking," "authentic learning," "hands-on" experiences, "authentic assessment," and many other of today's best known pedagogical terms and concepts are rooted in Dewey's adaptation of developmentalism. Other recent (but now less popular) forms of developmentalism are the "third force" and "humanistic" psychologies on which the educational innovations of the nineteen sixties and seventies were based (Weber, 1972).

A variety of other popular practices are less explicitly developmentalist but they share developmentalism's premises about the goodness of the natural--a characteristic that is key to their acceptance by the educational mainstream. Well known examples include the "whole language" and "language experience" approaches to reading (Altwerger, Edelsky & Flores, 1987), the closely related "emergent literacy" view of reading (Teal & Sulzby, 1987), and the "cognitive apprenticeship" approach to instruction (Brown, Collins, and Duguid, 1989). Stahl and Miller's (1989) discussion of whole language and language experience reading instruction highlights its appeal as a "natural" mode of instruction: "The goal of both approaches is to bring children into literacy in a 'natural' way [italics added], by bridging the gap between children's own language competencies and written language" (p. 88).

## Developmentalism: The Term and Its Referents

Although Stone (1991, 1993a, 1994) seems to have originated the use of "developmentalism" in reference to the doctrine discussed herein, similar terms have been used to denote developmentally informed educational practice. Sprinthall and Sprinthall (1987) used the term "developmentalists" in reference to educators who base their practices on developmental considerations. A similar term-- "philosophic-developmentalist"-- was used by Lawrence Kohlberg and Rochelle Mayer (1972) in reference to the views of John Dewey (1859-1952) and Jean Piaget (1896-1980). Dewey's and Piaget's views were termed "interactionist" and those of Jean Jacques Rousseau (1712-1778), "maturationist." In contrast to these precedents, developmentalism as used by Stone (1991, 1993a, 1994) refers to a broad doctrine that presumes "natural" ontogenesis to be optimal. Such a presumption is common to both maturationist and interactionist views of development; and it is implicit in Dewey, Piaget, Rousseau, and the others here termed developmentalists. As the term is used here, the "ism" in developmentalism is the uncontested assumption that the "natural" course of development, however conceived in theory, is the optimal possibility. It is an obscure but vital form of romantic naturalism--one thoroughly embedded in the American culture.

Stated broadly, developmentalism is the view of age- related social, emotional, and cognitive change that regards the optimal progression to be a fragile result of native tendencies emerging in a world congenial to their presumed wholesome nature. It emphasizes (a) the sufficiency of a natural inclination to learning, (b) the dangers of interference with native characteristics and proclivities, and (c) the desirability of learning experiences that emulate those thought to occur naturally. Social, emotional, and cognitive attributes that may be the unrecognized result of teacher and parent intervention are presumed by developmentalism to be manifestations of nature's normal trajectory. Man, his social contrivances, and indeed, civilization are seen as distinct from nature; and deliberate efforts to alter the course of child development are suspected of interfering with optimal developmental outcomes.

Developmentalism assumes that the developmental directions issuing from the child's native tendencies and characteristics are optimal because they are a part of "nature." Although their concepts of development differed, Rousseau, Dewey, Piaget, and all other developmentalists share this premise. For Rousseau, nature was God's work untainted by human influence. In his view, the optimal developmental progression was simply the emergence of native tendencies and characteristics unfettered and unspoiled by society. By contrast, Dewey and Piaget considered the child's tendencies and characteristics to be the product of Darwinian evolution. Native tendencies and characteristics were desirable because they had survived the process of natural selection. Unlike Rousseau, Dewey and Piaget held that the optimal progression depended not only on successful maturation but on a natural process of interaction wherein the native characteristics selected-for by evolution were enhanced by the naturally occurring experiences to which they were fitted (Kohlberg & Mayer, 1972). Thus originated Dewey's emphasis on authentic educational experience. Evolution equipped humans to learn by solving problems, therefore learning in the context of problem solving was optimal. Although Rousseau's development was more exclusively a matter of maturation, he too treated social and educational influences as having the ability to either facilitate and nurture, or to corrupt and misdirect the optimal progression to which nature was postulated to tend.

### A Brief History of Developmentalism

Developmentalism's historic foundations go well beyond the writings of Rousseau, Dewey, and Piaget. Pedagogical theorists such as Johann Bernard Basedow (1724-1790), Johann Heinrich Pestalozzi (1746-1827), Georg Wilhelm Friedrich Hegel (1770-1831), Friedrich Froebel (1782-1852), Herbert Spencer (1820-1903), William James (1842-1910), and G. Stanley Hall (1844-1924) are the best known proponents of the past 200 years. In general, their views were premised on either the maturation-only or the maturation/environmental- interaction schemes of development.

The ascendance of developmentalism in America may be related to an early belief about education as a cause of madness. According to Makari (1993), Rousseau's "education naturelle" was presaged by the writings of John Locke in 1691 and Giambattista Vico in 1709. Vico believed that children develop through a series of immutable phases and he condemned educational practices not in harmony with the "natural" progression. He considered abstract Cartesian thought to be particularly harmful. Vico's supposition that that which appears to be unnatural is apt to harmful has been echoed repeatedly even to the present day. Proponents of "developmentally appropriate" teaching practice, for example, believe that the use of incentives with young children are likely to be damaging.

Vico's belief was accepted within American psychiatry from its earliest years, and it persisted in the professional literature well into the late eighteen hundreds (Makari, 1993). The public and professional acceptance of such thinking as enlightened and informed clearly would have lent credibility to the criticisms of formalist teaching methods voiced by Dewey, James, and others. Also it would have bolstered the acceptance of the developmentalist schooling methods imported from Europe throughout the era.

### Rousseau and European Developmentalists

Rousseau argued that all that comes from the hand of the Creator must be good; and in doing so, he substituted a doctrine of original goodness for that of original sin. He believed that formal schooling was not only unnecessary (because children tend naturally to learn) but that it harms students by violating their natural propensities (Green, 1955). Classically premised on a romanticist faith in nature, Rousseau's Emile was a critique of educational practice in his day.

Hegel embellished Rousseau's theme and described child development as a process of unfoldment toward a state of natural perfection (Bigge & Hunt, 1962). Basedow, Pestalozzi, and Froebel each articulated their unique vision of schooling based on Rousseau's and Hegel's concepts (Rusk, 1965). In each case, their conceptual framework required schooling to be fitted to the child in the interest of preserving the goodness inherent in nature, and in each case they were received by the European public as a welcome alternative to the often harsh teaching methods of the day. Teachers of the era typically were retired drill sergeants and their methods were adaptations of military training (Riegel, 1972).

### Herbert Spencer and William James

Spencer and James similarly argued that education must be fitted to the child but their ideas were premised on an evolutionary model of nature (Cremin, 1964). The vision of natural perfection suggested by evolutionary theory differed from that of Rousseau but the ideal of education in harmony with natural perfection again was perpetuated. Optimal educational results were those that arose from fulfillment of nature's inherent order--an order shaped by the workings of evolution. Although Spencer and James both relied on an evolutionary premise, their thinking diverged as to the relationship between the natural order and desirable educational outcomes. Spencer conceived of education as subordinate to and, ideally, accommodated to the broader evolutionary process. He held that men were "infinitely more creatures of history than its creators" (Cremin, 1964, p. 93). Thus educational practice fitted to nature's dictates was the arrangement most conducive to optimal enhancement of the species. In contrast, James conceived of the human mind as having an active role in shaping the natural order and; more than Spencer, Rousseau, or Dewey, he believed that teachers should instill good (i.e., adaptive) habits.

James differed in other important ways from Dewey and other developmentalists. In contrast to Dewey, James conceived of educational outcomes as specific observable behavior change, not as a broad gaged and intangible intellectual growth. Also in contrast to Dewey and most other developmentalists, James believed that learned habits could serve to inhibit or overcome unfavorable natural tendencies. Thus he was he was not especially critical of recitation and the older "formalist" educational methods, and neither did he expect all learning to be motivated by a genuine personal interest. In James's words, the belief that learning should be motivated only by interest was "soft pedagogy" (James, 1899/1924, p. 109).

As to the relationship between human development and learning, James held that evolution had endowed humans with naturally "ripening" instincts and native interests to which successful teaching should be fitted. Unlike Dewey and other developmentally informed theorists, however, he did not insist on adherence to nature's ripening process or on an approximation of nature's interaction patterns as the optimal means of educating. Rather James' Talks to Teachers (1899/1924) offered practical recommendations that could be implemented largely without reference to developmental considerations. Thus in spite of his attention to human development as an educational consideration, James, unlike Dewey, did not greatly contribute to the restrictive orthodoxy that is developmentalism.

### G. Stanley Hall and Arnold Gesell

G. Stanley Hall may have been the individual most responsible for infusing the American educational tradition with the maturation-only version of developmentalism (Strickland & Burgess, 1965). Hall believed that quality teaching was that which was fitted to what he termed a "saltatory" pattern of development--a pattern he believed to have been dictated by human evolutionary history (Hall, 1907).

Hall's views are among the most explicitly developmentalist in the history of American education; and although his "general psychonomic law" (ontogeny recapitulates phylogeny) was eventually rejected, his concept of improving the educational process through the study of child development became a mainstay educational orthodoxy (McCullers, 1969). In his essay "The Ideal School as Based on Child Study," Hall argued that contrary to accepted Western educational practice, the school should be fitted to the child rather than the child fitted to the school. Teachers, he believed,

> . . . should strive first of all to keep out of nature's way, and to prevent harm, and should merit the proud title of defenders of the rights and happiness of children. They should feel profoundly that childhood, as it comes fresh from the hand of God, is not corrupt, but illustrates the survival of the most consummate thing in the world; they should be convinced that there is nothing else so worthy of love, reverence, and service as the body and soul of the growing child. (cited in Cremin, 1964, p. 103).

In his definitive account of progressive education, Cremin (1964, p. 104) argues that the popularization of Hall's "pediocentric" view was "truly Copernician" because it shifted the "burden of proof" for learning from the student to the school. Coming at a time when compulsory education was becoming widespread, its impact on American education was enormous and continues to be felt.

The aim of improving the educational process through child study was further popularized by Hall's student Arnold Gesell. Although not widely read today, Gesell's developmental concepts are consistent with popularly held views of early childhood development (cited in Bigge & Hunt, 1962):

> As with a plant, so with a child. His mind grows by natural stages. A child creeps before he walks, sits before he stands, cries before he laughs, babbles before he talks, draws a circle before he draws a square, lies before he tells the truth, and is selfish before he is altruistic. Such sequences are part of the order of Nature. . . . Every child, therefore, has a unique pattern of growth, but that pattern is a variant of a basic ground plan. (p. 166)

## John Dewey and Progressive Education

John Dewey is another developmentalist who did not rely on a formally stated developmental sequence. Instead, Dewey believed that evolution had equipped man with characteristics fitted to certain types of naturally occurring experiences and that the learning that emerges as the individual encounters these experiences is optimal. Quality teaching was, therefore, the practice of fitting educational experiences to the emerging characteristics and proclivities of the child for the purpose of optimizing "growth." Optimal development was both driven by maturation and nurtured by experience. In contrast to Rousseau, Dewey did not consider maturation sufficient to guide the process. Instead, he was frequently critical of progressive educators who followed Rousseau's maturational precepts, referring to their ". . . idealizing of childhood [as] . . . lazy indulgence" (cited in Axtelle & Burnette, 1970, p. 260).

Also contrary to popular belief, Dewey conceived of school as a structured experience in which teachers would ingeniously arrange student encounters with personally meaningful problems--problems which, if well chosen, would instigate self-directed learning experiences (Dewey, 1916/1963). The teacher's actions, however, were intended as a means of facilitating or enhancing a spontaneous learning process, not as a means of unnaturally or artificially inducing a preconceived outcome. In Dewey's words, the only proper aim of education is "growth" (Dewey, 1916/1963):

> Since growth is the characteristic of life, education is all one with growing; it has no end beyond itself. The criterion of the value of school education is the extent in which it creates a desire for continued growth and supplies means for making the desire effective in fact. (p. 53)

Dewey argued that the right sort of experience would instigate "reflective" thinking and thereby move the student toward a meaningful and individually defined form of knowing. The problem solving experience was, in his view, nature's way of teaching--the way in which the species had been equipped for learning by virtue of natural selection. Dewey's prescriptions for teaching were designed to emulate nature's process.

Because he believed that true understanding was personalized, Dewey held that educational aims could not be dictated by any agent external to the student (Dewey, 1916/1963, 1938/1963; Feldman, 1934/1968). For this reason, Dewey's concepts severely limited the ability of teachers to insure that students acquire any preconceived understanding or knowledge. Education was a process intended to enhance the student's reflective powers. That subject-matter which a student learned incidental to the educational process was the only important or expected kind of formal educational achievement--a view clearly at odds with traditional expectations for schooling and with the concept of teacher accountability for specific academic accomplishments. An individual's familiarity with the knowledge and insights gleaned by intellectual forebearers was of secondary importance in Dewey's thinking.

Dewey's departure from traditional expectations for schooling was tied to his reliance on an evolutionary model of nature (Boydston, 1970). He believed that progressive schooling would produce varied outcomes; that the outcomes most advantageous to society would be selected for; and that society would be bettered by the process. Although he opposed preconceived outcomes as the aim of schooling, his faith in human rationality led him to expect that students would arrive at commonly held truths as a result of their personal explorations.

A similarly founded departure from conventional expectations for schooling--Dewey's emphasis on student interest as the sole legitimate source of student motivation- -led to practical difficulties with his approach. Because student interests might be far removed from conventional academic pursuits, the time, effort, and resources necessary to elicit their emergence was destined to collide with economic reality. The cost-effectiveness of schooling was not a major consideration in Dewey's time. Neither was the availability of meaningful occupational opportunities for students whose natural thirst for learning was significantly delayed. Thus in spite of his pragmatic orientation, neither Dewey nor his followers seemed to appreciate the pedagogic and economic inefficiencies that would result as growing children became immersed in a world increasingly dominated by competing attractions.

As to reliance on formal knowledge of human development, Dewey called for teachers to be guided by the emergence of the individual student but to be informed by known developmental considerations (1916/1963):

> The method of [knowing and learning exhibited by an individual student] . . . will vary from that of another (and properly vary) as his original instinctive capacities vary, as his past experiences and his preferences vary. Those who have already studied these matters are in possession of information which will help teachers in understanding the responses different pupils make, and help them in guiding these responses to greater efficiency. Child-study, psychology, and a knowledge of the social environment supplement the personal acquaintance gained by the teacher. But methods remain the personal concern, approach, and attack of an individual, and no catalogue can ever exhaust their diversity of form and tint. (p. 173)

In essence, the student's "needs" were to guide the selection and sequencing of educational experiences. Accordingly, Dewey's curriculum was comprised of the subject matter and experiences that fit the unique pursuits of the individual. Knowledge of formal subject matter was purely incidental to the educational process (Dewey, 1938/1963).

The fact of Dewey's long and prestigious career combined with the extensive influence of the progressive education movement resulted in Dewey's principles and its inherent developmentalism becoming a very potent educational orthodoxy. Cremin (1964) notes that by the late nineteen forties and early fifties, the language and concepts of progressive education were no longer thought of as representing a particular educational view. Rather they were simply considered good and sensible educational practice. For a period of fifty or so years following World War I, both the U. S. Office of Education and the National Education Association disseminated educational recommendations based on progressive principles as "best practices." Today, teaching practices inspired by Dewey's concepts continue to attract adherents despite discouraging empirical findings. The attempt to improve student achievement by matching teaching styles with learning styles and investigations of attribute-treatment interactions are examples of research that fail to support Dewey's recommendations for teaching (Slavin, 1991).

Within teacher education, progressivists were extremely influential. William Heard Kilpatrick held the senior chair in social foundations of education at Teachers College, Columbia University from 1918 to 1938. During that time he is said to have taught 35,000 teachers (Cremin, 1964). Thus even though progressive education per se eventually fell into disrepute, its concepts and jargon were so thoroughly established as "conventional wisdom" that the reasonableness and intuitive appeal of all subsequent educational theorizing was largely governed by its compatibility with progressive concepts--concepts that for the most part embodied one or another version of developmentalism.

## Neoprogressive Theorists

Subsequent to progressive education's demise in the late nineteen fifties, a number of neoprogressive psychological theories, all possessing a strong developmentalist bent, gained widespread popularity within the teaching profession (Weber, 1972). Exemplars include Lawrence Frank, Daniel Prescott, Carl Rogers, Arthur Combs, Abraham Maslow, A. S. Neill, and Erik Erickson--all of whom viewed central aim of education as a broad gauged personal development. Although their theoretical foundations and emphases diverged from those of progressive education, (for example, the liberation of human potential, the enhancement of self-esteem, the achievement of self-actualization, etc.), their recommendations for teachers were plainly congruent with progressive education's focus on facilitation of naturally developing tendencies and processes. Other theorists emphasized narrower facets of development but they too were entirely compatible with developmentalism and progressive education (Weber, 1972). These include Paul Torrence who focused on the development of intellectual creativity and Lawrence Kohlberg who articulated a moral development progression based on Piaget's general framework.

Of particular relevance to present day educational practice are the neoprogressive accounts of cognitive development that became popular in the late nineteen sixties and early seventies. Jerome Bruner and, especially, Jean Piaget are the best known exemplars in this area; and both are essentially compatible with Dewey, particularly in their emphasis of a natural, i.e., personal discovery, type of learning experience.

### Jean Piaget and Lev Vygotsky

As earlier noted, Kohlberg and Mayer (1972) identified both Piaget and Dewey as exponents of "philosophic- developmentalism"--a view that holds intellectual growth to be the only defensible aim of education. Piaget's theory was grounded in his extensive observations of his three children and in a host of more systematic investigations undertaken subsequently. By training a biologist, Piaget described what seemed to be a biologically shaped sequence of person/ environment interaction--one he believed necessary to the emergence of individual intelligence. Thus, in contrast to the commonsensical and anecdotal accounts of intellectual development offered by Dewey, Piaget's work provided educators an elaborate theoretical edifice based on legitimate scientific observation.

The Russian psychologist Vygotsky (1987), a contemporary of Piaget, similarly conceived of a biologically shaped developmental progression but with an important differences in emphasis. In contrast to Piaget, Vygotsky argued that learning as a result of sociocultural experiences played a far greater role in the emergence of mature thinking and behavior. The influence of experience on behavior, however, was limited by a biologically governed zone of proximal development. Of the two theorists, Piaget was far better known and thus exerted far greater influence on educational practice.

Given the credibility of his findings, Piaget's educational recommendations were taken as substantially more authoritative and convincing than those of Rousseau, Dewey, and the others. Yet, despite its merits, Piaget's theorizing did not escape the preconceptions of its predecessors. As had Dewey and Rousseau, Piaget surveyed that which he took to be the naturally occurring developmental progression and presumed it optimal. Thus his conclusions--ones buttressed by impressive theoretical and empirical refinements-- conferred a predictable and welcome affirmation of developmentalist beliefs.

Piaget's educational recommendations were intended to preserve "natural" experiences and to facilitate that which is unique to the individual. According to Kohlberg and Mayer (1972) they include:

> . . . (1) attention to the child's mode or style of thought, i.e., stage; (2) match of stimulation to that stage, e.g., exposure to modes of reasoning one stage above the child's own; (3) arousal, among children, of genuine cognitive and social conflict and disagreement about problematic situations (in contrast to traditional education which has stressed adult "right answers" and has reinforced "behaving well"); and (4) exposure to stimuli toward which the child can be active, in which assimilatory response to the stimulus-situation is associated with "natural" feedback. (p. 462 )

Although the empirical underpinnings of Piaget's framework have been undermined by subsequent research (Siegler, 1991) and his theory significantly revised (Case, 1991), Piaget's thinking remains highly influential with mainstream educators. Its recent educational expression is the increasingly well known "constructivism" (Brooks & Brooks, 1993); and as with virtually all popular educational doctrines, its acceptance by the educational mainstream reflects its compatibility with Dewey and developmentalism. Overton (1972) acknowledges the mutually supportive relationship between Piagetian developmental concepts and Dewey. In essence, Dewey enabled popularization of Piaget, and Piaget has provided a seemingly unassailable rationale for Dewey's educational prescriptions:

> . . . Piaget's functional position contributes primarily to educational foundations and methods. The implications of his major emphasis upon activity echo progressive education's assertions of intrinsic motivation, self-direction, and freedom of the learner. The detailed analysis of the nature of the activities involved in adaptation stresses the significance of discovery-oriented methods in which the teacher actively participates by presenting appropriate materials and setting appropriate problems over methods of rote drill, training, or enriched environments. Above all, there is the point shared with progressive education that learning and development occur through the experience of the child's actively confronting his social and physical world. (Overton, 1972, p. 113-114)

Thus the theoretical and empirical expressions of present day (mainly Piagetian) developmentalism may not be Dewey's but its conclusions about educational practice are largely the same (Reschly & Sabers, 1974).

Although today viewed principally as guide to teaching at the primary school level, developmentalism serves as a conceptual foundation for educational practice at all levels (Clark & Starr, 1991; Sprinthall & Sprinthall, 1987; Squire, 1972; Wlodkowski, 1986). At the preschool and K-3 levels, the "developmentally appropriate instruction" concept has so thoroughly penetrated educational thinking that it is included in the "America 2000" statement of national educational goals (U.S. Department of Education [USDOE], 1991); it is acknowledged in the school reform principles formulated by business leaders (Committee for Economic Development, 1991); and it is explicitly cited in school reform legislation (Kentucky Education Reform Act, 1990; Stone, 1993).

### Developmentalism's Restrictions on Teaching and Parenting

Developmentalism's effect on educational reform must be understood in the context of its influence on teaching, parenting, and socialization as a whole. As the now popular African proverb suggests, "it takes a village to raise a child," thus the influence developmentalism's strictures and recommendations on the actions of both parents and teachers are critical to schooling outcomes.

In general, developmentalist guidance has encouraged parents and teachers to be less assertive and to afford children greater freedom. In particular, it has encouraged lessened parent insistence on study and effort in school and on mature and responsible behavior generally. Parents are given to believe that in a developmentally accommodative world, frustration and delayed gratification are to be minimized while immediate success and satisfaction are to be maximized. For example, an NEA publication by Wlodkowski (1986), discourages teachers' from insisting on results:

> We need to look more at the process and performance of our students and less at the more narrow and self-defeating emphasis of product or acquisition. If a student is responding with enthusiasm and interest, she/he will probably learn, but often without a neat, continuous, daily progress line. To lose our students'

> excitement and involvement for lack of immediate learning is not only a waste of effort but also a danger to the ultimate goal of any teacher--a student who is on the road to becoming a lifelong learner. (p. 16)

The National Association for the Education of Young Children (NAEYC) is more specific. Its policy statement on "developmentally appropriate practice" identifies that the following actions to be inappropriate (Bredekamp, 1988):

> The teacher's role is to correct errors and make sure the child knows the right answer in all subject areas. Teachers reward children for correct answers with stickers or privileges, praise them in front of the group, and hold them up as examples. (p. 76)

Broadly speaking, developmentalism and its restrictions on teaching practice argue against intervention and, instead, favor the kind of premissiveness found in the child-rearing recommendations of Dr. Benjamin Spock (1976) and others (Brazelton, 1974; Gessell & Ilg, 1943; Warner & Rosenberg, 1976). In truth, Spock, et al and the educational developmentalists rely on many of the same theoretical foundations.

Developmentalism suggests that both teacher and parent expectations for behavior or achievement must be subordinated to concerns about optimal development. Rather than seek to shape the child to social or academic norms, developmentally informed teachers and parents are deemed responsible for affording experiences and opportunities that are compatible with the child's current proclivities. That such experiences will result in effort and achievement commensurate with individual potential is simply taken for granted. Clark and Starr (1991, p.37) exemplify this view in their textbook on secondary and middle school teaching methods: "Because learning is developmental, it follows that one learns better when one is ready to learn." Bigge and Hunt's (1962, p. 377) text is more explicit: "A young person is ready to learn something when he has achieved sufficient physiological maturation and experiential background so that he not only can learn but wants to."

Whatever the measurable impact of developmentally informed teaching and parenting on the course of child development (a remarkably little examined topic), its immediate impact on teacher and parent attempts to instruct and discipline are entirely foreseeable. Developmentalism gives rise to a disabling hesitancy and uncertainty about how or whether adults should attempt to influence children. It strongly suggests the possibility of harm, but it offers no clear guidance as to a safe and effective course of action. It requires an estimation of a child's developmental status as a prerequisite to action yet it offers no workable means of ascertaining that status.

The requirement of correctly inferring individual development presents a substantial obstacle to the application of developmental theory. The prototypic studies of human development by Gessell (1940, 1943, 1946), Gesell, Ilg, and Ames (1956), and McGraw (1945/1969) tracked physical and motor development--both low inference constructs. The indicators of development--height, weight, number of teeth, number of steps, etc.--were visible and readily quantifiable. By contrast, the phases of social, emotional, and cognitive development to which developmentally appropriate teaching and parenting must be fitted are high inference constructs, i.e., ones said to be manifested by complex patterns of behavior. The inherent observational problem is evident in Piaget's concept of intelligence (Furth & Wachs, 1975):

> For Piaget, intelligence is constructive and creative; in fact, development of intelligence is but the gradual creation of new mechanisms of thinking. It is creation because it is not the discovery or the copy of anything that is physically present. Classes and probability cannot be found in the physical world. They are concepts constructed creatively by human intelligence

> and cannot be handed down by means of language or other symbols. (pp. 25-26)

To add to the imprecision and uncertainty of the required inference, Piaget's theory holds that the relationship between current behavior and developmental status is neither fixed nor self-evident and that the underlying developmental progression is characterized by spurts, lulls, and uneven dispersion across the various behavioral, emotional, and intellectual domains. Again in reference to Piaget (Furth & Wachs, 1975):

> This variability takes three forms, each of which is contrary to a normative ideal. First, different individuals differ on the same task and much more than an IQ mentality would have us believe. . . . A second type of variability is found within a certain individual (intraindividual variability) as he performs on a variety of different tasks [tasks requiring the same underlying intellectual capability]. . . . A third type of variability is observed both within the same individual and on the same task. In other words, the performance of a child fluctuates from day to day--an entirely normal phenomenon that all of us experience. . . . Recognition and acceptance of this variability is particularly important in the case of mechanisms of thinking which develop gradually and almost imperceptibly [italics added]. (pp. 28-29)

In addition to their ambiguity, estimates of developmental status are inherently conservative and restrictive of adult action. Conceptually, current levels of intellectual performance, effort, maturity, achievement, and other indicators can understate but not exceed present levels of development. For example, a child whose reasoning is concrete operational may exhibit skills indicative of the earlier preoperational level but they would never misleadingly exhibit skills appropriate to the more mature formal operations level. Thus assessments of development based on a child's current behavior may underestimate but not overestimate present developmental status.

Given that developmentally appropriate teaching and parenting must be fitted to the child's current developmental status, and given that efforts to exhort or otherwise induce advancement beyond the child's developmentally governed potentialities are considered risky at best, teachers and parents are given to understand that expecting too little is a much better choice than expecting too much. From a developmentalist perspective, if opportunity and conditions conducive to developmental advancement have been maximized, the developmentally guided teacher or parent has done all that can safely be done.

In effect, developmentalism discourages teachers and parents from asserting expectations or otherwise acting to induce more mature behavior. Even in the face of noticeable deficiencies or problematic conduct, the developmentally appropriate course of action is that which is congenial to the child's apparent developmental status, i.e., his or her present behavior and inclinations. Continuing lack of advancement in spite of suitable facilitating conditions is taken to reflect delayed emergence of developmentally governed potentialities, not ineffective teaching or parenting.

**Personal, Social, and Cultural Implications**

The implications of such a perspective are far-reaching and they may be relevant to the well known concerns about the waning influence of homes and schools. In a world that affords few immediate incentives for responsible and constructive behavior, children whose teachers and parents are captivated by developmentalism may be significantly disadvantaged: They are too little influenced by those adults who have the greatest interest in their well being. To the extent that teachers, parents, and other socially ordained influences are withheld, "default contingencies" (John Eshleman, personal communication, February 26, 1993)-- i.e., influences arranged by peers, by the entertainment and recreation industries, etc.--are empowered.

Not only does developmentalism appear to undermine teacher and parent assertiveness, the view of children inherent in developmentalism may be negatively linked to the "growth" of maturity, character, and a sense of personal responsibility. Rather than encouraging parents to treat children and youth as individuals responsible for their own behavior, developmentalism encourages tolerance and acceptance of immaturity, irresponsibility, and failure. And given the belief that mature and responsible behavior simply emerges if properly facilitated, the child who fails to exhibit expected social and academic progress is excused as a victim of adverse circumstances--a rationale for individual shortcomings that has become a cultural archetype (Birnbaum, 1991).

The influence of developmentalism and its philosophic foundation, romantic naturalism, may extend far beyond teaching and parenting practices. For example, the growth of so called "anti-science" (Holton, 1993; Kurtz, 1993) and of certain forms of environmentalism seem to be linked to the same romantic assumptions about the wholesomeness of nature that are integral to developmentalism. Over a 75 year period developmentalism has been a prominent feature of educational practice, and from this venue, it has had opportunity to thoroughly infuse the American culture. The degree to which popular thought in America may have been influenced by romanticist leanings within the public schools, however, is well beyond the present analysis.

### Implications for Schoolwork

Learning of the kind sought by schools inevitably requires very substantial commitments of student time and effort (Tomlinson, 1992). Developmentalism, however, discourages teachers from any attempt to directly induce it. Instead, developmentalism requires that teachers endeavor to produce "learning in ways that are stimulating yet minimally obtrusive, challenging yet requiring only comfortable levels of exertion" (Stone, 1994, p. 65). An anomaly becomes apparent (Stone, 1994):

> . . . schools [are encouraged] to spare neither effort nor resources in fitting instruction to students while expecting little from them in return. Student inattention and apathy are met with herculean efforts to stimulate interest and enthusiasm. Deficient outcomes are countered by reducing expectations to the level of whatever the student seems willing to do. Even the practice of [motivating students by] affording . . . accurate feedback about accomplishments is deemed questionable because of its purported detrimental effect on intrinsic motivation and self esteem.
> . . . recurrent failure to attain even minimal achievement is accepted as lamentable but unavoidable and treated accordingly. In short, developmentalism requires only the teacher to work, not the student. (p. 62)

In essence, developmentalism leads to schools in which attendance is compulsory but study is not. Students are expected to make an effort only if they feel interested and enthused. Study is expected to be more like fun than work. If students waste time and educational opportunity because they find schoolwork boring, their behavior is not merely tolerated, it is understood and excused as the product of insufficiently stimulating instruction, i.e., instruction that fails to facilitate the emergence of the postulated ideal.

In the end, teachers are burdened with an unattainable expectation. They, their employers, and the public are encouraged to believe that if a teacher is sufficiently creative and ingenious in harnessing each individual student's potentialities, expected learning outcomes will emerge in a way that the student will experience as spontaneous, natural, and comfortable. It is an ideal founded wholly on developmentalist supposition but it has come to define good teaching.

Developmentalism's ideal of taking the work out of schoolwork may be responsible not only for poor work habits and attitudes beyond the classroom--a problem widely noted by employers (Mandel, Melcher, Yang & McNamee, 1995; Survey, 1991). So long as study and effort are expected only if the student feels so inclined, the self discipline necessary to putting school "work before pleasure" is largely omitted from the academic regimen. Instead of a work ethic, students are given to expect significant accomplishments with minimal effort (Shine, 1993).

### Educationally Appropriate Practice

A vital distinction must be drawn between developmentally appropriate instruction and educationally appropriate instruction, i.e., those teaching practices that accommodate teaching to the learner without regard to the hypothetical constraints posed by developmental theory. Developmentally appropriate instruction (a.k.a. developmentally appropriate practice) seeks to optimize the development of the "whole child" (Johnson & Johnson, 1992) irrespective of academic norms. It is a "learner centered" (a.k.a. "student centered" or "child centered") approach to teaching (Darling- Hammond, Griffin and Wise, 1992) meaning that the teaching process is constrained by developmental considerations but the product is open ended. It is an approach that rejects both expectations for accomplishment based on curricular benchmarks or peer referenced norms as well as any "artificial" means of insuring that they materialize.

In contrast, "educationally appropriate" instruction (Stone, 1994) seeks to meet recognized standards and to otherwise maximize academic achievement. Both developmentally appropriate and educationally appropriate instruction rely on present levels of demonstrated performance as a starting point for instruction and both seek to optimize intellectual advancement. Educationally appropriate teaching (or practice), however, does not treat present performance as a marker for a child's developmental limits. It is "learning centered" in the sense that observed performance, not presumed developmental limitations, guides academic advancement. Although sensitive to student comfort with teaching practice, educationally appropriate practice holds achievement, not developmental suitability, to be its top priority and neither does it presume high expectations or teacher insistence on effort to be developmentally hazardous.

In conclusion, developmentalism appears to discourage teacher and parent intervention while simultaneously promoting the belief that academic achievement and responsible behavior will spontaneously emerge if only given time and facilitating conditions. Contrary to developmentalist expectations, however, it may be that awaiting the emergence of wholesome behavior is an open invitation to default contingencies and the growth of unfavorable habits--ones that might have been precluded by the acquisition of appropriate patterns. By the time the realities of such deficits and/or inappropriate conduct make the need for action inarguable, remediation is likely to be more difficult. Well ingrained patterns of faulty behavior must first be eliminated before constructive alternatives can be established--a situation all too familiar to special educators and school psychologists.

### The Developmentalist Neglect of Experimentally Vindicated Teaching Practices

Developmentalism influences teacher acceptance of experimentally demonstrated teaching practices in much the same way it impacts teaching and parenting generally. It argues against intervention on the grounds that it is likely to detract from the more optimal outcome that presumably will emerge when natural developmental processes are permitted to run their course.

### Some Neglected Methodologies

Over the last thirty years, a variety of experimentally vindicated teaching methods have been developed and disseminated only to be ignored or discarded in favor of less well tested practices that better fit developmental thinking. Mastery learning and Personalized System of Instruction may be the best known examples (Kulik, Kulik, & Bangert-Drowns, 1990). Direct

Instruction (Becker & Carnine, 1980)--also known as DISTAR (Kim, Berger, & Kratochvil, 1972) and as "systematic instruction" (Slavin, 1994)--is another. Direct Instruction is little used despite having been as thoroughly validated and field tested as any methodology in the history of education (Watkins, 1988). These and a large group of structured and sequenced teaching methodologies termed "explicit teaching" (Rosenshine, 1986) are among the most clear instances of experimentally supported approaches to teaching that have failed to gain widespread acceptance and/or have been abandoned.

Programmed instruction (Skinner, 1958) is another example of an abandoned methodology and one that uniquely appears to demonstrate how developmentalism's hold on the teaching profession influences teaching practices in public schools. Despite its initial acceptance and evident promise, K-12 educators rejected programmed instruction in favor of less structured, more naturalistic, "real-world," "hands-on" approaches (Skinner, 1986). However, among educators less influenced by developmentalism, i.e., private sector business and industrial trainers, military trainers, designers of computer-based instruction, etc., it remained well established (Ellson, 1986; Vargas & Vargas, 1992).

Many of the experimentally validated methodologies are behavioral because behavioral approaches to teaching and learning are derived from the experimental analysis of behavior. However, mastery learning (Bloom, 1976) and the "explicit teaching" methodologies discussed by Rosenshine (1986) are not behavioral and the same can be said for most of the "productive" methodologies discussed by Ellson (1986) and Walberg (1990, 1992). Ellson (1986) listed seventy-five studies of teaching methods all of which report learning effects that are at least twice as great as control comparisons. Most of these methods were popular at one time but none are in widespread use today. Walberg (1990, 1992) summarized the results of nearly 8000 studies that point to the efficacy of a brief list of powerful and teacher- alterable classroom interventions, most of which are supported by experimental evidence. High expectations for effort and achievement is one, the use of incentives is another. In general, the neglected methodologies identified by Walberg and Ellson are structured and teacher directed; they aim to instill preconceived academic and intellectual outcomes; and most of them employ practice, feedback, and incentives.

## Developmentally Inspired Concerns, Reservations, and Objections

Teaching methods textbooks and other sources of recommendations about teaching practice seem to sanction the disuse of experimentally vindicated methodologies either by giving them little or no attention or by discussing them in the context of various concerns, objections, and reservations (Jacobsen, Eggen, & Kauchak, 1993; Ornstein, 1992; Wlodkowski, 1986). These remarks are especially noticeable when contrasted to the uncritical treatment given developmentally compatible methodologies. Typical cautions and criticisms involve claims that the experimentally vindicated methods are insufficiently individualized (Armstrong, 1980), too artificial and mechanical (Bailey, 1991), excessively reliant on extrinsic motivation (Kohn, 1993a, 1993b), suited only to lower forms of learning (Ornstein, 1992), or simply boring (Henson, 1993; Lemlech, 1994). Virtually all of these reservations and objections are premised on a developmentalist view of learning.

Developmentalists hold that adherence to that which is developmentally appropriate is more important than educational achievement thus they favor educational experiences that are well accepted by students over those that are known to produce results. In the developmentalist view, teachers should seek methods that produce results but they should select them only from among those methods that maximize student satisfaction. Judged by priorities so ordered, experimentally vindicated teaching methodologies are suspect at best because they are built around the notion that learning is the primary consideration. If the authors of methods textbooks were to suggest that teachers should prefer methodologies that have been experimentally vindicated, they would be in disagreement with developmentalist doctrine, i.e., with the view that student satisfaction is primary and learning secondary. The same consideration applies to teacher expectations for student effort and achievement. Developmentalism suggests that teachers should expect a commitment to schoolwork that is commensurate with the student's lifestyle and developmentally determined inclinations, not with external and artificial requirements that are based on arbitrary or socially derived academic standards.

In effect, developmentalism requires experimentally vindicated practices not only to be attractive, interesting, and engaging, it obliges them to overcome the belief that they are likely to be risky or harmful, i.e. that they interfere in unknown or unsuspected ways with a virtually boundless range of developmental considerations (Elkind, 1981). The test of usefulness to which demonstrably effective interventions are subjected is not one of observed cost and benefit compared to the observed cost and benefit of an existing alternative, it is one that entails suspected hidden cost versus the perfection that hypothetically emerges in the absence of human interference.

For example, when "whole language" proponents express concern about skill-sequence approaches to reading (Goodman & Goodman, 1979), they worry that the interest in reading that otherwise naturally emerges might be lessened. Criticisms of drill, corrective feedback, and the use of incentives are typically founded on the same argument. If, however, nature is permitted the opportunity (i.e., a "developmentally appropriate" opportunity) to work its effects, developmentalists assume that the expected skills and interest will emerge and without exposure to the hazards inherent in intervention (Clark & Starr, 1991; Lemlech, 1994; Jacobsen, Eggen, & Kauchak, 1993; Stone, 1995).

### The Alleged Threat to Intrinsic Motivation.

Some developmentally inspired reservations about experimentally vindicated methodologies are based on more than theoretical extrapolations. For example, the concerns about reductions in intrinsic motivation due to positive reinforcement reported by Deci & Ryan (1985), Lepper, Greene, & Nisbett (1973), and Schwartz (1990) appear to be supported by credible empirical findings. Even these claims, however, seem to have been exaggerated without challenge perhaps as a result of developmentalism's enormous influence within the educational community.

For the past seventy-five or so years, the teaching profession has idealized learning that is motivated by interest as the only "true" learning. Led by Dewey (1916/1963; 1938/1963), the mainstream teaching profession has held that such "intrinsic" or naturally occurring interest will express itself provided that the student is confronted with a sufficiently meaningful or relevant or lifelike problem. Thus teaching that relies on extrinsic sources of motivation is, according to Dewey's concept, inherently poor teaching, i.e., insufficiently creative, innovative, and stimulating, and its use of extrinsic incentives a concession to faulty educational practice. The widespread acceptance of Dewey's developmentally informed vision seems likely to have contributed to the positive reception given the reports of Deci, Ryan, Lepper, et al. and, more recently, to Kohn's (1993a, 1993b) wholesale derogation of positive reinforcement, incentives, rewards, and competition.

The technical foundations of these reports, however, have been the subject of scholarly disagreement, and the exaggerated nature of their claims has become evident in the recent meta-analysis by Cameron and Pierce (1994). Reviewing the literature from 1971 to the present, they conclude that the empirical findings with respect to intrinsic motivation simply do not warrant exclusion of incentives from the classroom.

One other telling observation may be made about Kohn's (1993a, 1993b) criticisms. Positive reinforcement and other extrinsic sources of motivation have been successfully employed by school psychologists, special educators, and teachers of remedial and "at risk" students for many years (Hallahan, Kauffman, & Lloyd, 1985; Hammill & Bartel, 1990). Apparently that evidence has been overlooked or discounted. Perhaps such applications are considered exempt from developmentalist strictures because students to whom they are applied have acknowledged developmental imperfections.

Despite their success, however, interventions that are known to benefit the disabled are not entirely immune from criticism. For example, there is ongoing debate among early childhood special educators regarding "early intervention" versus "developmentally appropriate practice." Again, the question is one of whether successful experimentally founded intervention strategies are producing some subtle but as-yet- unnoticed developmental harm (Carta, Schwartz, Atwater & McConnell, 1991; Johnson & Johnson, 1992).

### The Alleged Inattention to Thinking.

Of the developmentally inspired concerns pertaining to experimentally vindicated teaching methods, their alleged neglect of student thinking is, by far, the most frequent criticism (Armstrong & Savage, 1994; Callahan, Clark, & Kellough, 1992; Clark & Starr, 1991; Henson, 1993; Jacobsen, Eggen, & Kauchak, 1993; Kim & Kellough, 1995; Lemlech, 1994; Ornstein, 1992; Sheperd & Ragan, 1992). These concerns and the current pedagogical emphasis on cognitive processes, higher-order intellectual skills, critical thinking, reflective thinking, etc., again, reflect Dewey's (1916/1963) view of learning:

> The sole direct path to enduring improvement in the methods of instruction and learning consists in centering upon the conditions which exact, promote, and test thinking. Thinking is the method of intelligent learning, of learning that employs and rewards the mind. (p. 153)

The same can be said of the present day emphasis on hands-on, authentic, real-world learning experiences as a means of facilitating learning:

> Only by wrestling with the conditions of . . . [a] problem at first hand, seeking and finding his own way out, does . . . [the student] think. When the parent or teacher has provided the conditions which stimulate thinking and has taken a sympathetic attitude toward the activities of the learner by entering into a common or conjoint experience, all has been done which a second party can do to instigate learning. The rest lies with the one directly concerned. (Dewey, 1916/1963, p. 160)

Both Dewey (1916/1963) and Piaget (Siegler, 1991) considered human learning capabilities the product of evolutionary demands for intellectual adaptation to the natural world. Formal knowledge and skills were held to be important only to the extent that they were integrated with applications to problem solving. If natural circumstances required humans to learn and employ knowledge in the context of problem solving, Dewey reasoned that schools would optimize learning by doing the same. Thus in Dewey's scheme of education, thinking in service of problem solving is primary to education and acquisition of formal knowledge and competencies is secondary and incidental.

What Dewey may not have adequately considered is that traits evolved under one set of conditions can prove useful under other conditions and in service of entirely different ends. For example, human hands were not initially selected- for because of their usefulness in writing or musical performance but they subsequently served that purpose. Analogously, the ability to acquire and retain knowledge may have been selected-for under conditions where knowledge was wholly contextualized, yet today the same ability can be usefully employed to acquire knowledge that is partly or wholly decontextualized.

Given the advantages that industrial and technological cultures appear to derive from formal instruction afforded in a classroom setting, it seems evident that a profitable use has been found for the human ability to acquire factual, abstract, and decontextualized knowledge and that acquisition of such knowledge is a useful prerequisite to real-world, problem solving experiences. In fact, it would seem that schooling in societies which make use of the formal knowledge cumulated from the experiences of innumerable ancestors would necessarily entail a substantial amount of decontextualized learning. Thus the achievement of preconceived objectives through experimentally vindicated teaching methodologies may afford socially, economically, and pedagogically advantageous gains in educational efficiency despite its inconsistency with the ideals inherent in Dewey, Piaget, and other popular theorists.

### Why Non-experimental Research is Better Accepted

In contrast to the skepticism typically encountered by experimentally founded interventions, teaching practices informed by studies of naturally occurring social and educational processes are relatively well received by the educational community. Even if not adapted to developmental considerations, such practices do not suggest artificially imposed alterations of "natural" conditions. Thus if peer interaction processes or certain teacher or student characteristics are found to be correlated with student achievement, teachers can be safely encouraged to take advantage of these "natural" (and presumably causal) relationships by creatively interpreting and selectively employing them as developmental considerations permit. Studies of relationships between educational outcomes and student learning styles (Dunn, Beaudrey, & Klavas, 1989; Shipman & Shipman, 1985) are a good example. The recent surge of recommendations favoring greater sensitivity to multicultural diversity in the schools also seem founded on this type of research (Boykin, 1986; Thompson, Entwisle, Alexander, & Sundius, 1992). In each case, these studies encourage teachers to shape instruction to the preferences and inclinations of the student in order to enhance achievement to the extent that student proclivities will permit.

Unfortunately, of course, the causal inferences suggested by descriptive and correlational studies can be grossly misleading and their misinterpretation has lead to some of the most egregious instances of faulty teaching practice. The attempt to improve learning by boosting self- esteem is a prime example (Scheirer & Kraut, 1979).

**The Incompatibility of Developmental and Experimental Views**

Given the nature of the developmentalist view, experimentally demonstrated teaching practices are bound to invite a great degree of skepticism. The object of experimental research is to demonstrate the impact of an independent variable as an agent of change. Contrary to such an objective, developmentalism requires that social, emotional, and cognitive change emerge, not as an effect induced by an external agent, but as an independent expression of the student. Thus experimentally tested methodologies are automatically considered suspect if not outrightly objectionable depending on which developmental limitations are presumed applicable. In effect, developmentalist doctrine discourages reliance on the most important and most credible research educators have at their disposal (Bloom, 1980 as cited in Gage & Berliner, 1992; Cook & Campbell, 1979).

Because they claim an applicability that never seems adequately tempered by developmental considerations, experimentally validated methods tend to encounter an impassable gauntlet of questions and reservations. In a reference to Walberg's (1984) report of generalizable, robust, and teacher- alterable influences on learning, Ralph Tyler (1984) expressed the forlorn hope that the (developmentalist) notion that each student and each circumstance is so unique that it can only be understood (i.e., effectively taught) by a teacher deeply immersed in the situation would be dispelled.

Armstrong (1980) raised the same issue in discussing teacher demand for educational research:

> Given the nature of undergraduate teacher preparation programs and the cultural milieux of large numbers of schools, many teachers have come to believe that teaching is more art than science. Exposed to much talk about "individual differences" and "unique characteristics" of every classroom, many view teaching and teaching problems as situation-specific. Through their training and interactions with many colleagues, large numbers of teachers are more predisposed to acknowledge the differences than the commonalties characterizing the human condition. Consequently, many teachers suspect any generalized statements about human behavior. This orientation prompts many to doubt the value of educational research efforts that, by design,

seek generalizable knowledge [italics added]. (p. 59)

The restrictions on effective practice posed by developmentalism have largely precluded many otherwise credible attempts to improve education through applications of science. The contrast between the degree of scientifically founded progress in medicine versus that found in education attests this conclusion. To a large extent, medical science has benefitted man by employing scientifically informed means of intervening in nature. The artificial creation of immunities through the use of "unnatural" and invasive vaccination is an historic example. In contrast, educational improvements on "natural" patterns and processes of learning have been severely restricted by a doctrine of developmentalism. Instead of using experimentally validated teaching methods, teachers have been encouraged to emulate nature and thereby preserve the perfection assumed to exist in natural developmental processes.

## Conclusion

Developmentalism presumes typical patterns and processes of social, emotional, and cognitive change to be optimal because they are "natural." It fails to recognize the extent to which valued social, emotional, and cognitive attributes may be induced and sustained (not merely facilitated) by the purposeful actions of teachers and parents. Indeed, it seems to underestimate the importance of civilizing influences generally. By default, developmentalism ascribes the positive effects of unrecognized environmental influences to "natural" processes and argues that attempts to alter their effects are likely to be harmful.

Present day developmentalism frames the process of socialization and, specifically, that of teaching as one of influencing the child in such a way as to avoid disruption of a postulated optimal outcome. It transforms teaching from an endeavor straightforwardly concerned with achievement to a search for naturalistic conditions that will fit the learner's tendencies in a way that permits the unfettered and, therefore presumably optimal, emergence of intellectual growth. Developmentalism assumes that teaching which deviates from this general prescription is, at best, naive and, at worst, dangerous and destructive of the learner's best interests. Thus teaching practices uninformed by developmental considerations are persistently rejected by the teaching profession regardless of demonstrated educational effectiveness and otherwise wholesome impact--a pervasive and powerful but largely unrecognized restriction on scientifically founded educational improvement.

## References

Alexander, R., Rose, J., & Woodhead, C. (1992). Curricular organization and classroom practice in primary schools, a discussion paper. London (SW1P 3BT), United Kingdom: Department of Education and Science.

Altwerger, B., Edelsky, C., & Flores, B. M. (1987). Whole language: What's new? The Reading Teacher, 41, 144-154.

American Legislative Exchange Council [ALEC], (1994). Report card on American education 1994. Washington, DC: Author.

Armstrong, D. G. (1980). Researchers' incentives and the dearth of practical research results. Peabody Journal of Education, 58, 55-60.

Armstrong, D. G., & Savage, T. V. (1994). Secondary education, an introduction (4th ed.). New York: Macmillan.

Axtelle, G. E., & Burnette, J. R. (1970). Dewey on education and schooling. In J. A. Boydston (Ed.), Guide to the works of John Dewey (pp. 257-305). Carbondale, IL: Southern Illinois University Press.

Bailey, J. S. (1991). Marketing behavior analysis requires different talk. Journal of Applied Behavior Analysis, 24, 445-448.

Becker, W. & Carnine, D. (1980). Direct instruction: An effective approach for educational intervention with the disadvantaged and low performers. In B. Lahey and A.Kazdin (Eds.), Advances in child clinical psychology. New York: Plenum.

Berliner, D. C., & Biddle, B. J. (1995). The manufactured crisis. New York: Addison-Wesley.

Berliner, D. & Biddle, B. (1996, February). Making molehills out of molehills: Reply to Lawrence Stedman's review of The Manufactured Crisis. Education Policy Analysis Archives, 4(3). http://olam.ed.asu.edu/epaa/

Bigge, M. L., & Hunt, M. P. (1962). Psychological foundations of education: An introduction to human development and learning. New York: Harper & Row.

Binder, C., & Watkins, C. L. (1989). Promoting effective instructional methods: Solutions to America's educational crisis. Future Choices, 1 (3) 33-39.

Birnbaum, J. (1991, August 12). Crybabies: Eternal victims. Time, pp. 16-18.

Bloom, B. S. (1976). Human characteristics and school learning. New York: McGraw-Hill.

Boydston, J. A. (Ed.). (1970). Guide to the works of John Dewey. Carbondale, IL: Southern Illinois University Press.

Boykin, A. W. (1986). The triple quandary and the schooling of Afro-American children. In U. Neisser (Ed.), The school achievement of minority children. Hillsdale, NJ: Erlbaum.

Bracy, G. (1996). International comparisons and the condition of American education. Educational Researcher, 25(1), 5-11.

Brazelton, T. B. (1974). Toddlers and parents. New York: Delacorte Press.

Bredekamp, S. (1988). NAEYC position statement on developmentally appropriate practice in the primary grades, serving 5 through 8 year olds. Young Children, January, pp. 64-84.

Brooks, J. G., & Brooks, M. G. (1993). In search of understanding: The case for constructivist classrooms.

Alexandria, VA: Association for Supervision and Curriculum Development.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18, 32-42.

Bruner, J. S. (1966). Toward a theory of instruction. New York: Norton.

Callahan, J. F., Clark, L. H., & Kellough, R. D. (1992).Teaching in the middle and secondary schools (4th ed.).New York: Macmillan.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: a meta-analysis. Review of Educational Research, 64(3), 363-423.

Campbell, D. & Stanley, J. (1963). Experimental and quasi-experimental designs for research. Skokie, IL: Rand McNally.

Carnine, D. (1993, December 8). Facts over fads.Education Week, p. 40.

Carnine, D. (1995, May 3). Is innovation always good? Education Week, p. 40.

Carta, J. J., Schwartz, I. S., Atwater, J. B., & McConnell, S. R. (1991). Developmentally appropriate practice: Appraising its usefulness for young children with disabilities. Topics in Early Childhood Special Education, 11 (1), 1-20.

Case, R. (Ed.). (1991). The mind's staircase: Stages in the development of human intelligence. Hillsdale, NJ: Erlbaum.

Clark, L. H., & Starr, I. S. (1991). Secondary and middle school teaching methods (6th ed.). New York: Macmillan.

Committee for Economic Development. (1991). The unfinished agenda: A new vision for child development and education (A statement by the Research and Policy Committee for the Committee for Economic Development). New York: Author.

Cook, T. D., & Campbell, D. T. (1979). Quasi- experimentation, design & analysis issues for field settings. Boston: Houghton Mifflin.

Cremin, L. A. (1964). The transformation of the school, progressivism in American education 1876-1957. New York: Vintage Books.

Darling-Hammond, L., Griffin, G., Wise, A. (1992).Excellence in teacher education: Helping teachers develop learner-centered schools. Washington, DC: National Education Association.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York: Plenum.

Dewey, J. (1916/1963). Democracy and education. New York: The Macmillan Company.

Dewey, J. (1938/1963). Experience and education. New York: Collier Books.

Dunn, R., Beaudrey, J. S., & Klavas, A. (1989). Survey of research on learning styles. Educational Leadership, 46(6), pp. 50-58.

Elkind, D. (1981). The hurried child. Newton, MA: Addison- Wesley.

Ellson, D. (1986, October). Improving productivity in teaching. Phi Delta Kappan, pp. 111-124.

El-Nemr, M. A. (1980). A meta-analysis of the outcomes of teaching biology as inquiry (Doctoral dissertation, University of Colorado, 1979). Dissertation Abstracts International, 40, 5813A. (University Microfilms International No. 80-11274)

Feldman, W. T. (1934/1968). The philosophy of John Dewey, a critical analysis. New York: Greenwood Press.

Freedman, J. (1993). Failing grades, Canadian schooling in a global economy. Red Deer, Alberta, Canada: Society for Advancing Educational Research.

Furth, H. G. & Wachs, H. (1975). Thinking goes to school, Piaget's theory in practice. New York: Oxford University Press.

Gage, N. L. & Berliner, D. C. (1992). Educational psychology. Boston: Houghton Mifflin.

Gesell, A. & Ilg, F. L. (1940). The first five years of life; a guide to the study of the preschool child. New York: Harper.

Gesell, A. & Ilg, F. L. (1943). Infant and child in the culture of today. New York: Harper and

Row.

Gesell, A. & Ilg, F. L. (1946). The child from five to ten.New York: Harper.

Gesell, A., Ilg, F. L., & Ames, L. B. (1956). Youth: The years from ten to sixteen. New York: Harper.

Giacomia, R. M., & Hedges, L. V. (1982). Identifying features of effective open education. Review of Educational Research, 52, 579-602.

Goodman, Y., & Goodman, K. S. (1979). Learning to read is natural. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1, pp. 137- 154).Hillsdale, NJ: Erlbaum.

Green, F. C. (1955). Jean Jacques Rousseau: A critical study of his life and writings. New York: Cambridge University Press.

Guskey, T. R., & Pigott, T. D. (1988). Research on group- based mastery learning programs: A meta- analysis.Journal of Educational Research. 81, 197-216.

Hall, G. S. (1907). Youth, its education, regimen, and hygiene. New York: Appleton.

Hallahan, D. P., Kauffman, J. M., & Lloyd, J. W. (1985).Introduction to learning disabilities (2nd ed.).Englewood Cliffs, NJ: Prentice Hall.

Hammill, D. D. & Bartel, N. R. (1990). Teaching students with learning and behavior problems (5th ed.). Boston: Allyn and Bacon.

Heinemann. (1995). Books for teachers K-8, 1995.(Catalogue available from Heinemann-Boyton/Cook, College Department, 361 Hanover St., Portsmouth, NH 03801)

Henson, K. T. (1993). Methods and strategies for teaching in secondary schools (2nd ed.). New York: Longman.

Hetzel, D. C., Rasher, S. P, Butcher, L., & Walberg, H. J.(1980). A quantative synthesis of the effects of open education. American Educational Research Association.

(ERIC Document Reproduction Service No. ED 191 902) Holton, G. (1993). Science and anti-science. Cambridge, MA: Harvard University Press.

Jacobsen, D., Eggen, P., & Kauchak, D. (1993). Methods for teaching, a skills approach (4th ed.). New York: Merrill.

James, W. (1899/1924). Talks to teachers on psychology: and to students on some of life's ideals. New York: Henry Holt and Company.

Johnson, J. E., & Johnson, K. M. (1992). Clarifying the developmental perspective in response to Carta, Schwartz, Atwater, and McConnell. Topics in Early Childhood Special Education, 12(4), 439-457.

Kantrowitz, B., & Wingert, P. (1993, November). Failing the most gifted students. Newsweek, p. 67.

Kentucky Education Reform Act of 1990, KRS }} 156.010 (1990)

Kim, E. C. & Kellough, R. D. (1995). A resource guide for secondary school teaching, planning for competence (6th ed.). New Jersey: Prentice Hall.

Kim, Y., Berger, B. J., & Kratochvil, D. W. (1972). Distar Instructional System (Product Development Report No. 14).Palo Alto, CA: American Institute for Research.

Kirsch, I. S., Jungblut, A., Jenkins, L., & Kolstad, A.

(1993, September). Adult Literacy in America (executive summary). Washington DC: U. S. Department of Education.

Kohlberg, L., & Mayer, R. (1972). Development as the aim of education. Harvard Educational Review, 42, 449- 496.

Kohn, A. (1993a). Punished by rewards: the trouble with gold stars, incentive plans, A's, praise, and other bribes. New York: Houghton Mifflin.

Kohn, A. (1993b). Rewards versus learning: A response to Paul Chance. Phi Delta Kappan, pp. 783-787.

Kulik, C. -L., Kulik, J. A., & Bangert-Drowns, R. L. (1990).Effectiveness of mastery learning programs: A meta- analysis. Review of Educational Research, 60(2), pp. 265- 299.

Kurtz, P. (1994, March 12). The growth of anti- science.The Skeptical Inquirer, 18(3), p. 255.

Lemlech, J. (1994). Curriculum and instructional methods for the elementary and middle school (3rd ed.). New York: Macmillan College.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973).Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis.Journal of Personality and Social Psychology, 28, 129-137.

Lysakowski, R. S., & Walberg, H. J. (1980). Classroom reinforcement. Evaluation in Education, 4, 115-116.

Lysakowski, R. S., & Walberg, H. J. (1981). Classroom reinforcement and learning: A quantative synthesis. Journal of Educational Research, 75, 69-77.

Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and feedback: A quantative synthesis. American Educational Research Journal, 19(2), 559-578.

Madamba, S. R. (1981). Meta-analysis on the effects of open and traditional schooling on the teaching-learning of reading (Doctoral dissertation. University of California, Los Angeles, 1980). Dissertation Abstracts International, 41, 3508A. (University Microfilms International No. 81- 02856)

McCullers, J. C. (1969). G. Stanley Hall's conception of mental development and some indicators of its influence on developmental psychology. American Psychologist, 24, 1109-1114.

Makari, G. J. (1993). Educated insane: A nineteenth- century psychiatric paradigm. Journal of the History of the Behavioral Sciences, 29, 8-21.

Mandel, M. J., Melcher, R. A., Yang, D. J., & McNamee, M. (1995, April 17). Will schools ever get better? Business Week, pp. 18-21.

Marshall, J. (1993, December). Why Johnny can't teach. Reason, pp. 31-35.

May, H. (1959). The end of American innocence. New York: Alfred A. Knopf, Inc.

McGraw, M. B. (1945/1969). The neuromuscular maturation of the human infant. New York: Hafner Publishing Company.

National Education Association. (1995). If teacher training is your interest. (NEA Professional Library Publications catalogue available from the National Education Association, 1201 16th. St. N.W., Washington, DC 20036- 3290)

Ornstein, A. C. (1992). Secondary and middle school teaching methods. New York: Harper Collins.

Organization for Economic Co-operation and Development [OECD], (1994). OECD economic surveys 1993-1994: United States.

Paris, France: Author (OECD Publications, 2 rue Andre- Pascal, 75775 Paris CEDEX 16).

Overton, W. F. (1972). Piaget's theory of intellectual development and progressive education. In J. R. Squire (Ed.). A new look at progressive education (pp. 88-115). Washington, DC: Association for Supervision and Curriculum Development.

Peterson, P. L. (1980). Open versus traditional classrooms.Evaluation in Education, 4, 58-60.

Ravitch, D. (1983). The troubled crusade: American education 1945-1980. New York: Harper-Collins.

Reschly, D., & Sabers, D. (1974). Open education: Have we been here before? Phi Delta Kappan, 55, (10), 675-677.

Riegel, K. (1972). Influence of economic and political ideologies on the development of developmental psychology. Psychological Bulletin, 78(2), 129-141.

Riley, R. W. (1993, April). Statement by Richard W. Riley, U. S. secretary of education on the national assessment of educational progress 1992 mathematics report card for the nation and the states. Washington, DC: U. S. Department of Education.

Rosenshine, B. V. (1986). Synthesis of research on explicit teaching. Educational Leadership, 43(7), 60-69.

Rusk, R. R. (1965). The doctrines of the great educators. Revised and enlarged third edition. New York: St. Martins Press.

Scheirer, M. A., & Kraut, R. E. (1979). Increasing educational achievement via self concept change. Review of Educational Research, 49(1), 131-150.

Schwartz, B. (1990). The creation and destruction of value. American Psychologist, 45, 7-15.

Sheperd, G. D., & Ragan, W. B. (1992). Modern elementary curriculum (7th ed.). Fort Worth: Harcourt Brace Jovanovich College.

Shipman, S., & Shipman, V. (1985). Cognitive styles: Some conceptual, methodological and applied issues. In E. W.Gordon (Ed.), Review of Educational Research, Vol. 12. Washington, DC: American Educational Research Association.

Shine, B. (1993). Let's be realistic about what teachers can accomplish. (Available from Committee to Elect Shine for Governor, 433 East Center Street, Kingsport, TN 37660)

Siegler, R. S. (1991). Children's thinking (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Skinner, B. F. (1958). Teaching machines. Science, 128, 969-977.

Skinner, B. F. (1986, October). Programmed instruction revisited. Phi Delta Kappan, pp. 103-110.

Slavin, R. E. (1994). Educational psychology, theory and practice. Boston: Allyn and Bacon.

Slavin, R. E. (1991). Educational Psychology. Englewood Cliffs, NJ: Prentice Hall.

Spock, B. (1976). Baby and child care (Rev. Ed.). New York: Pocket Books.

Sprinthall, N. A., & Sprinthall, R. C. (1987). Educational psychology: A developmental approach (4th ed.). New York: Random House.

Squire, J. R. (Ed.). (1972). A new look at progressive education. Washington, DC: Association for Supervision and Curriculum Development.

Stahl, S. A., & Miller, P. D. (1989). Whole language and language experience approaches for beginning reading: A quantitative research synthesis. Review of Educational Research, 59, 87-116.

Stanovich, K. E. (1992). How to think straight about psychology (3rd. ed.). New York: Harper Collins.

Stanovich, K. E. (1993). Romance and reality. The Reading Teacher, 47(4), pp. 280-291.

Stedman, L. C. (1996, January). The achievement crisis is real. [Review of The Manufactured Crisis]. Education Policy Analysis Archives, 4(1). http://olam.ed.asu.edu/epaa/

Stedman, L. C. (1996, April). Respecting the evidence: The achievement crisis remains real. Education Policy Analysis Archives, 4(7). http:// olam.ed.asu.edu/epaa/

Stone, J. E. (1991). Developmentalism: A standing impediment to the design of the "New American School." Network News & Views, 10 (12), 1-3.

Stone, J. E. (1993a, August). Behavioral technology in a developmentalist context. In S. Axelrod (Chair), Educational technology - concepts and applications. An invited symposium held at the 101st Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Stone, J. E. (1993b). An open letter to the Oregon state governing bodies. Effective School Practices, 12(2), 3-6.

Stone, J. E. (1994). Developmentalism's impediments to school reform: Three recommendations for overcoming them. In Ralph Gardner III, Diane M. Sainato, John O. Cooper, Timothy E. Heron, William L. Heward, John W. Eshleman, & Teresa A. Grossi (Eds.), Behavior analysis in education: Focus on measurably superior instruction. Pacific Grove, CA: Brooks/Cole.

Stone, J. E., (1995, May). Motivational recommendations found in popular teaching methods textbooks. In C. McMahon (Chair), Conceptual and Theoretical Issues in Education and Educational Reform. A paper presented at the 21st Annual Convention of the Association for Behavior Analysis, Washington, DC.

Strickland, C. E. & Burgess, C. (Eds.). (1965). Health, growth, and heredity, G. Stanley Hall on natural education. New York: Teachers College Press.

Survey: Fewer skilled workers. (1991, December 10). Johnson City Press, p. 14.

Teal, W. & Sulzby, E. (1987). Emergent literacy: Writing and reading. Norwood, N. J.: Ablex, 1987.

Tomlinson, T. (1992). Hard work and high expectations: Motivating students to learn (Superintendent of Documents, P. O. Box 371954, Pittsburgh, PA 15250-7954; order #065-00496-8). Washington, DC: Office of Research and Improvement, U. S. Department of Education.

Thompson, M. S., Entwisle, D. R., Alexander, K. L., & Sundius, M. J. (1992). The influence of family composition on children's conformity to the student role.American Educational Research Journal, 29(2), 405-424.

Tyler, R. (1984). A guide to educational trouble- shooting.Educational Leadership, 41, 27-30.

U.S. Department of Education. (1991). America 2000, an education strategy (revised). Washington, DC: Author.

Vargas, E. A. & Vargas, J. S. (1992). Programmed instruction and teaching machines. In R. P. West & L. A. Hamerlynck (Eds.), Designs for excellence in education: The legacy of B. F. Skinner (limited edition, pp. 33-69). Longmont, CO: Sopris West, Inc.

Vygotsky, L. S. (1987). The collected works of L. S. Vygotsky (translated by Rieber and Carton). New York: Plenum.

Walberg, H. J. (1990, February). Productive teaching and instruction: Assessing the knowledge base. Phi Delta Kappan, pp. 470-478.

Walberg, H. J. (1992). The knowledge base for teacher productivity. International Journal of Educational Reform, 1 (1), 1-10.

Walberg, H. J. (1984). Improving the productivity of America's schools. Educational Leadership, 41, 19-27.

Wang, M. C., Reynolds, M. C., & Walberg, H. J. (1987). Handbook of special education. New York: Pergamon.

Warner, S. L. & Rosenberg, E. B. (1976). Your child learns naturally. New York: Doubleday.

Watkins, C. L. (1988, July). Project follow through: A story of the identification and neglect of effective instruction. Youth Policy, pp. 7-11.

Weber, E. (1972). Conceptions of child growth and learning. In J. R. Squire (Ed.). A new look at progressive education (pp. 16-43). Washington, DC: Association for Supervision and Curriculum Development.

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? Educational Researcher, 21(5), 18-24.

Wlodkowski, R. J. (1986). Motivation and teaching: A practical guide. Washington DC: National Education Association.

## About the Author

J. E. Stone is a professor in the Department of Human Development and Learning at East

Tennessee State University. An Ed. D. graduate of the University of Florida, he is a licensed educational psychologist and school psychologist. Since 1972, he has taught more than 10,000 classes in ETSU's College of Education. His primary scholarly interest is educational reform in both K-12 and higher education. Currently, he heads the Education Consumers Clearinghouse--an internet networking and information resource for parents and other consumers of education.

P. O. Box 70548, ETSU
Johnson City, TN 37614
(423) 439-4190
stonej@access.etsu-tn.edu

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letterto LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the*Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or sendan e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass,Glass@asu.edu or reach him at College of Education,Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*
Alan Davis
*adavis@castle.cudenver.edu*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
William Hunter
*hunter@acs.ucalgary.ca*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Dewayne Matthews
*dm@wiche.edu*
Les McLean
*lmclean@oise.on.ca*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Richard C. Richardson
*richard.richardson@asu.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*

Andrew Coulson
*andrewco@ix.netcom.com*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Ernest R. House
*ernie.house@colorado.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 9** **June 12, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES and Andrew J. Coulson.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# Markets Versus Monopolies in Education: The Historical Evidence

**Andrew Coulson**

*a_coulson@msn.com*

**Abstract:** A common point of contention among educators and economists is the likely effect a free market would have on modern education. Most supporters of public schooling maintain that the field would either be adversely affected by competition and choice, or that the effects would be insubstantial. Conversely, a significant number of critics argue that education, like all other human exchanges, would respond to market incentives with improved performance, increased attention to the needs of families, and greater innovation. Historical evidence is presented indicating that teachers and schools are indeed affected by the financial incentives of the systems in which they operate. In particular, the data show that economic pressures have forced schools in competitive markets to meet the needs of families, through methodological advancements and diversity in curriculum, while centralized bureaucratic systems have generally been coercive and pedagogically stagnant.

## Introduction

The debate over educational funding and administration is an old one. Writing to his friend Tacitus almost two thousand years ago, the Roman lawyer Pliny the Younger described his plan to establish a secondary school in his home town, but added that he had decided to pay only one third of the total cost.

> I would promise the whole amount were I not afraid that someday my gift might be abused for someone's selfish purposes, as I see happen in many places where teachers' salaries are paid from public funds. There is only one remedy to meet this evil: if the appointment of teachers is left entirely to the parents, and they are conscientious about making a wise choice through their obligation to contribute to the cost. (Pliny, 1969, p. 277-283)

Over the last decade, proposals for introducing a degree of parental choice and inter-school competition into education have abounded, particularly in the United States, the

United Kingdom, Australia, and New Zealand. In some cases, such plans are already in place. With few exceptions, though, current choice programs pose barriers to the entry of new schools and to the exit of unpopular ones, exclude religious and/or profit-making institutions, restrict admissions and staffing policies, and otherwise control the supply and demand for education. Though private schooling exists in most industrialized countries, there is only limited competition at the primary and secondary levels. The comparatively heavy burden of tuition, when compared to the "free" status of tax-supported schools, greatly limits the clientele for private education. This in turn keeps the density of private institutions to a much lower level than if government did not provide schools. As a result, there is no nation currently offering a truly free and competitive market in education.

## The Case Against

As market-inspired reform has gained in popularity, it has been subjected to a great deal of criticism. Attacks have been directed at the possible ill-effects of parental- choice, of for-profit schools, and of market systems as a whole. The most often heard argument against a market is that parents cannot be expected to make sound educational choices for their children, and must instead leave the key decisions to experts. A significant number of parents, it is assumed, would either fail to inform themselves about competing schools, or would base their choices on the "wrong" criteria. This contention has been directed at the population as a whole (Carnegie Foundation, 1992; Wells & Crain, 1992), and also at specific groups such as the poor or the poorly-educated (Payne, 1993; Levin, 1991; Kozol, 1992). A related criticism is that racial and economic isolation might be increased if families selected their schools based on race, ethnicity, or social status (Cookson, 1994; Kozol, 1992).

On the supply side, skeptics argue that for-profit schools with bold promises, flashy advertising, and special programs would lure customers away from academically superior institutions (Krashinsky, 1986). Murnane (1983), and others have noted the possibility of fraud in voucher systems, in which corrupt principals could offer kick-backs to parents who chose their institutions. Profit-making schools are also expected by some critics to reject difficult-to-educate children, e.g. those with disabilities or serious discipline problems. According to Shanker and Rosenberg (1992), these children would be more expensive to teach and hence would either be expelled more readily or refused admission entirely.

All these objections have in common the idea that education is fundamentally different from other human exchanges, and that as a result, the natural checks and balances of the market would fail to operate as they normally do. There is a second line of argument that takes the opposite position, namely, that an educational market would fail precisely because it would operate in the same way as other markets (Krashinsky, 1986). Education, so the argument goes, benefits not only the students and their families, but their fellow citizens as well. These indirect benefits are said to include social harmony, political stability, and a thriving economy. According to Levin (1991), public school systems are capable of producing the aforementioned benefits, while a competitive market of private schools could either not produce them at all, or do so only at prohibitive regulatory expense.

The remaining criticisms are based on the results of "limited choice" or "public school choice" programs, which place many restrictions on schools and families, and generally do not allow the participation of private or parochial schools. Smith and Meier (1995), for example, argue that since programs allowing parents to choose from among different public schools have failed to substantially increase student learning, the same should be expected of an unregulated market. The experience with heavily regulated parental choice in the Netherlands (Brown, 1992; Elmore, 1990) is also cited in arguments against the effectiveness of competition. In the United States, comparisons between existing public and private schools have led Cookson (1994) to conclude that a market would not improve education. The same author also reasons that since private schools have rarely been included in choice programs, there is insufficient evidence to support free market educational reform.

## The Case in Favor

Virtually all of the criticisms discussed above have been disputed by proponents of parental choice. Members of the minority groups assumed to be incompetent or uninterested in their children's education are foremost in defending their ability and prerogative to choose. State representative Polly Williams (1994), herself an African-American single parent, championed a private school choice plan in Milwaukee Wisconsin on the grounds that public schooling had failed the urban community and that competitive private provision offered a superior education. Similar arguments have been made by Native- American educator Ben Chavis (1994). Empirical studies have shown that poor parents with limited formal education, from Massachusetts (Fossey, 1994) to the mountain villages of Nepal (Pande, 1977), can and do choose schools on rational grounds (see also U.S. Dept. of Education, 1995; Martinez et al. 1994).

Arguments that racial segregation would increase under a free market have been challenged from two different perspectives. The late James Coleman (1990) observed that racial segregation within the American public school system was greater than that among private schools. So, while the percentage of African-American students in the public sector is greater than the percentage in the private sector, public schools are more likely to be all-white or all-black than their private counterparts. Opposing the very essence of the segregation claim are educators such as Derrick Bell (1987), who believe that the freedom to create separate schools for African Americans would be a boon rather than a hardship.

The assertion that private schools might defraud parents is commonly countered with the argument that such problems exist everywhere, including public schools. The cases of East St. Louis (Schmidt, 1995) and Washington D.C. are notorious examples. Rinehart and Lee (1991) note that a competitive market would at least exert pressure on a school to deal honestly and fairly with parents in order to maintain a healthy reputation, while the public monopoly offers educators no such incentive. Along the same lines, John Coons (1991) has observed that public schooling has not engendered the external benefits of social harmony and effective democracy assumed by its defenders. The American experience of Protestant bias in the education of immigrants at the turn of the century, as well as government-enforced racial segregation, are presented as evidence of this claim. Coons also contends that by removing the coercive element from school selection and allowing parents to choose for themselves, the goal of effective democracy would be strengthened.

To resolve the issue of difficult-to-educate children, Myron Lieberman (1991), investigated the current practices among private institutions. He found that rather than focusing on easy-to-educate students, the single largest group of for-profit schools actually serves the disabled. Studies have also suggested that urban private schools are able to maintain a higher level of discipline than their public counterparts with few if any admissions requirements, and only infrequent student expulsions (Blum, 1985).

For the supporter of free markets, objections based on public school choice programs are seen as misguided. To function effectively markets require significant competition, the lure of profit-making, and a minimum of restrictions on buyers and sellers. Few if any of these criteria hold among existing choice programs (OECD, 1994), and as a result it is argued that they cannot be expected to show any significant benefits (Lieberman, 1989).

The above rebuttals aside, the economic case for an educational market rests on two main presumptions: that monopoly control of education leads to coercion, indifference to the needs of families, and stagnation in the form and content of instruction, while competition and the profit motive would lead to greater quality and efficiency. The first case has been made at both national and school levels. While inflation-adjusted per-pupil spending in U.S. public schools tripled between 1959/60 and the present (U. S. Department of Education, 1993), test scores either held constant or declined (Sowell, 1993; Boaz, 1991). Comparisons between public school administrations and those of the private Catholic sector have shown the public bureaucracy to employ as many as thirty times the number of administrators per-pupil (Boaz, 1991). On a school by school basis, Eric Hanushek (1986; 1989) studied correlations between spending and student achievement only to find that the relationship was not statistically significant. Similar results have been reported by Childs & Shakeshaft (1986). Because of the absence of any truly competitive market in education, little direct contemporary evidence is available to demonstrate its effects on efficiency or achievement. In those cases where a limited degree of competition does exist, however, Hoffer et al. (1990), Borland and Howsen (1993), and others have found small but significant positive effects. Outside the field of education, the superiority of markets to monopolies is widely accepted, and Winston (1993) has demonstrated that reductions in regulation are generally associated with lower prices and better services for consumers, and even yield higher revenues for producers.

## The Present Work

As can be gleaned from the arguments cited above, the debate over a market in education has drawn almost entirely from the limited body of contemporary evidence. With the exception of E.G. West's (1994) analysis of 19th century England, the historical evidence regarding market vs. monopoly provision in education has been largely ignored. Education, however, is not a recent invention. Two and a half thousand years of schooling, from the informal to the regimented, from complete parental freedom to totalitarian domination, have preceded current practice. The study of educational history thus offers a wealth of insights into the effects of monetary incentives and centralized administration on the actions of parents and educators.

The next section looks at the educational experiences of four historical periods and places: classical Greece, Germany at the Reformation, England during the eighteenth and nineteenth centuries, and France after the Revolution. This selection is a more or less representative sample from a larger survey of the subject currently in progress. The most valuable lessons these histories have to teach us concern the relationship between school governance and school quality. In particular, they highlight the differences between markets and centralized bureaucratic school systems on three important measures of school performance: how well they respond to and satisfy the demands of parents and students (e.g. through innovation and diversity in curriculum), the degree to which they benefit their students directly (e.g. higher literacy, job/life skills), and their indirect benefits to the rest of society (e.g. thriving economy, social harmony).

## Educational Choice: Over Time and Around the World

### Greece

Formal education made perhaps its earliest appearance in China, well before the first millennium B.C., but the most suitable starting point to our study lies half a world away, in Greece. Unlike the uniform system of the Chinese, ancient Greek education developed along disparate and conflicting lines. This contrast, between parental freedom and state control, was best represented by the city-states of Athens and Sparta. By the fifth century B.C., schooling in both of these societies had become a general preparation for citizenship and adulthood, but the content and delivery of that preparation differed dramatically. It is with this organizational juxtaposition that we begin.

With the exception of requiring two years of mandatory military training, the government played little or no role in Athenian schooling. Socrates is said to have described the practice of the day as follows:

> When boys seem old enough to learn anything, their parents teach them whatever they themselves know that is likely to be useful to them; subjects which they think others better qualified to teach, they send them to school to learn, spending money upon this object. (Freeman, 1904)

Anyone who wished might open a school, setting whatever curriculum and tuition they deemed appropriate. The schools were operated as private enterprises, and so the subjects taught and fees charged were established by what parents wanted their children to learn, and how much they were willing to pay for that learning. Choosing a teacher was considered an important decision, and it was expected that a person would consult with friends and relatives, deliberating for several days on the matter (Plato, 1937). Competition to attract parents and students seems to have held costs to a relatively low level, since even the poorest families are thought to have sent their sons to school for a few years, despite the absence of state funding (Cole, 1960). It should be noted, however, that most girls and much of the slave population received little or no education in Athens, as in so many cultures up to modern times.

Schooling began at the age of six or seven, but wealthy parents likely sent their children to school earlier and kept them there for longer than did parents with limited means. This occurred not only because of the need to pay school fees, but also because poor and middle class families could not afford to support their children indefinitely, and so had to ensure that they learned a trade or craft through apprenticeship; an experience quite distinct from schooling. Even in this time-honored tradition, however, the Athenians were innovators. When a boy was apprenticed to a tradesman other than his father, his parents would draw up a statement indicating which skills they expected him to be taught and the tradesman received payment only if he provided the stipulated training (Freeman, 1904).

At the elementary level, Athenian parents sought three general categories of education for their children: gymnastics, music, and literacy. Competence in each of these areas was of great practical importance. Stamina, strength, and agility meant the difference between life and death at a time when wars were a constant threat, and every able bodied male citizen was expected to serve in the army. To understand the importance of musical instruction it must be remembered that Greek culture had been orally transmitted, largely in song, for centuries prior to the rise of Athens. Just as a grasp of reading and important works of literature are crucial to modern education, so was the knowledge and appreciation of epic poetry important in the 5th and 4th centuries B.C. Even as the social mores embodied in the oral tradition were codified and written down, the value Athenian citizens placed on music and poetry remained high. Writing began to rise in significance in the 5th century, as a tool for improving the political and judicial systems, for accurately recording the works of scientists, playwrights, and philosophers, and for making economic transactions more reliable. In the minds of the city's more philosophically oriented citizens, this combination of physical, musical, and intellectual development also satisfied an appreciation for harmony and balance in the human character.

While music and reading were probably taught in the same school, the study of gymnastics was carried out at a special location, called a palaestra, which consisted of changing rooms and an exercise field. The gymnastics teacher was expected to have an organized method of instruction which would improve stamina, strength, and agility, while keeping the risk of injury to a minimum. Physical trainers also seem to have to provided their students with nutritional advice (Plato, 1937). Children began their gymnastics training by performing aerobic exercise routines to build stamina and flexibility. As their bodies and skills developed, they were taught javelin and discus tossing, a variety of ball-games and other sports, and also wrestling and boxing.

At writing school, then as now, the child was first taught to recognize and write the letters of the alphabet. For the youngest children, this was done through song, and there is even a fragmentary play that survives from late in the 4th century B.C. in which the actors represented letters and formed syllables by pairing up with one another in the appropriate poses (Freeman, 1904). Once the child had learned his alphabet, he was taught to write on a folding wooden tablet covered with wax, into which he would etch letters with the pointed end of a stylus, and rub them out with the wide end. At first the writing teacher would lightly trace the letters, and the student would then scratch his pen over them in order to learn how to draw their shapes. Once he had mastered this step, the child would begin to write on his own (Plato, 1937).

As Athenian culture broadened and developed, the elementary school curriculum developed with it. More and more parents began to seek drawing and painting instruction for their children, and by Aristotle's time this had become a common option. Several generations later, these arts were considered a fourth core subject area, being studied by virtually all pupils (Marrou, 1965). Adaptation to the changing demands of parents and students was in fact a hallmark of Athenian education. Each step in the evolution of the society was matched by a corresponding change in the offerings of educators. The philosophers and scientists of the day were continually pushing forward the frontiers of human understanding, establishing in their wake a demand for a deeper and more comprehensive level of education. At the same time, the democratic franchise was extended to an ever larger segment of the population, and the powers of the assembly were growing apace. In order to win popular support in this vibrant democracy, it became necessary for would-be statesmen to not only offer compelling policies, but also to deliver them with clarity and elegance. Training in oratory was thus an important political asset. Together, the emerging educational demands of politics and science made higher-level teaching an economically viable endeavor. Athenians not only wanted to become better educated, they were willing to pay for it. This market niche was quickly filled by a new entrepreneurial class of teachers, known as sophists, anxious to earn a living from their scholarly pursuits.

At first, when the demand for higher-learning in any one community was still limited, the sophists traveled from city to city, holding forth on whatever topic they felt confident to teach, and for which there were eager pupils. When the flow of students had ebbed at a given location, they would once again resume their journey. Recruiting new pupils was always an important task for the sophists, since their livelihoods depended on it. The most common technique used to this end was the presentation of free public lectures in the town square, which allowed them to demonstrate their talents and whet the intellectual appetites of prospective students. Fortunately for the sophists, the spread of learning served not to diminish but rather to increase the demand for their services. As more and more people became better educated, the value of an education increased. It became necessary for anyone with hopes of public office or success in law or commerce to expand their educational horizons. This trend was not lost on elementary school masters who eventually began to diversify into the new secondary and higher education markets by offering advanced classes to adults and children over the age of fourteen. For many years, however, the bulk of higher-education was still carried out by the wandering professors.

While rhetoric and the sciences were the most common fields of study, the range of subjects taught by the sophists was astonishingly diverse. The curious student might choose from "mathematics (including arithmetic, geometry, and astronomy), grammar, etymology, geography, natural history [i.e. biology, horticulture, etc.], the laws of meter and rhythm, history..., politics, ethics, the criticism of religion, mnemonics, logic, tactics and strategy, music, drawing and painting, scientific athletics." (Freeman, 1904). Lectures were held in open spaces outdoors, in the homes of the teachers, and occasionally in buildings borrowed or leased for the purpose. There appear to have been no age restrictions on these lectures, and so any student both interested and capable of participating was permitted to do so.

Gradually, as the higher educational market matured, a few fixed schools were established in Athens. In addition to Plato's Academy and Aristotle's Lyceum, neither of which charged a fee due to the wealth and preferences of their founders, several for-profit secondary schools were in existence by the turn of the fourth century B.C. Only a few of these were sufficiently famous to come down to us by name, and of these the best known is the school of Isocrates. Contrary to Plato, Isocrates argued that knowledge without application was useless. He said, "I hold that man wise who can usually think out the best course to take and that man a philosopher who seeks to gain that insight."(Hamilton, 1957) Though reportedly too shy to become prominent in public life, Isocrates was extremely successful-both financially and by popular acclaim-in teaching the art of public speaking to others. This, coupled with his pragmatic lessons on applied philosophy and mathematics, attracted a significant body of students to his lectures. A greater number, it seems, than was to be found at the Academy. More remarkable though, and in a way more emphatically Athenian, was the school of Aspasia.

Defying the norms and prejudices of the day, this Milesian-born woman set up shop in Athens teaching philosophy and rhetoric, and unabashedly advocated the liberation and education of the city's women. According to Plato, her lectures attracted such towering figures as Socrates and Pericles, the latter of whom eventually became her lover and life-long companion. When asked of his ability to improvise a speech (in Plato's dialogue "Menexenus"), Socrates avowed that he was up to the task, and referring to Aspasia, added "I have an excellent mistress in the art of rhetoric-she who has made so many good speakers."(Plato, 1937) The philosopher goes on to suggest that one of the most famous speeches in ancient history, the funeral oration by Pericles, was actually written by her, and though there is little substantiation of this claim in the historical literature it certainly implies a healthy respect for her abilities on the part of Plato. Demonstrating the breadth of her appeal, Aspasia's school also attracted a large number of girls from well-to-do families, an emancipatory innovation that drew harsh criticism from many in the older generation (Durant, 1939). What is perhaps most significant about this case is the fact that, despite the intensely sexist climate of the city, the majority was not able to prevent Aspasia from opening her school and reaching out to the disenfranchised female population.

In stark contrast to the freedom and diversity of Athens, the central idea of Spartan society was that individuals and families should not be left to make their own decisions in matters of importance such as education, marriage, or employment. Instead, Spartans were called upon to second their own interests to the collective will of the people, as interpreted by their part aristocratic, part democratically- elected government. Supporting this sweeping centralization of authority was a monolithic educational apparatus run by the state, to which all citizens were compelled to send their sons (here again, the education of girls received less attention than that of boys). At age seven, all the male children were separated from their families and brought to live in school dormitories. The nature of their learning environment is well-captured by the terms used to describe them. A troop of boys was referred to as a "boua", the same word used for a herd of cattle, and from each herd, a dominant boy was chosen to act as herd-leader. With satisfying consistency, their head teacher was called "paidonomus", or boy-herdsman. This individual was chosen from the aristocracy, and granted the authority to train the boys, and to harshly discipline them if any failed to follow his instructions. In his efforts, he was assisted by two "floggers" armed with whips (Xenophon, 1988).

The children were administered an education consisting almost exclusively of sports, endurance training, and fighting. When questions were posed to the students, a prompt reply was expected, and those who failed to answer to the teacher's satisfaction were regarded as incompetent, and given a bite on the thumb or some similar punishment. Arithmetic is not mentioned as a part of the curriculum by any of Sparta's chroniclers, and few people could count beyond the smallest numbers. Students were perhaps introduced to letters, but certainly "no more than was necessary,"(Plutarch, 1988) and since books and written law were virtually non-existent in Sparta, this could not have been much at all. Isocrates did not hesitate to observe that the Spartans "have fallen so far behind our common culture and learning that they do not even try to instruct themselves in letters." (Isocrates, 1982) Speech and writing were further discouraged by an outright prohibition on learning rhetoric, the violation of which was a punishable offense (Sextus Empiricus, 1987). Educational innovation, whether it involved additions to the curriculum or the adoption of new techniques in the existing wrestling and military training, were strictly forbidden.

At dinner time boys were fed simple hearty meals, but were served deliberately small portions so that they would constantly be hungry if this were their only source of sustenance. To supplement this meager fare, children were encouraged to steal. Theft was in fact a central feature of Spartan education. The city's leaders believed that, if you want an army that thinks nothing of pillaging neighboring states, it is exceedingly helpful to have citizens accustomed to robbing their neighbors. While those caught stealing were severely punished, it was for failing to get away with the crime, rather than for attempting it in the first place. Skill in theft was considered a noble accomplishment, and, according to Isocrates, it paved the way to the highest political offices (Isocrates, 1982). Of course, students were encouraged to steal primarily from the subjugated peasant and slave populations rather than from other citizens.

By the time they had reached the age of eighteen, Spartan youths were tough, fit, ruthless, but also inexperienced. The missing element in their training was provided by an institution known as the "krypteia." Young men were gathered into bands and dispatched to the countryside where they would have to hunt and steal to survive. Their primary mission, however, was to attack their own peasant population whenever the opportunity arose, killing those who had the audacity to defend themselves. This savagery apparently seemed criminal even to the Spartans, for the elected officials would annually declare war on their own serfs, giving the bloodshed at least a veneer of legality.

Having described the different approaches to schooling in Athens and Sparta, we can look to the conditions of their people for a reflection of the effects of those systems. We cannot, of course, attribute all of the differences between Athenian and Spartan civilizations to their schools, but formal education clearly played an influential role.

To the classical Greeks, Athens was the "school of Hellas" and the "metropolis of wisdom." Of the three most influential philosophers in Western antiquity-Socrates, Plato, and Aristotle-the first two were Athenian citizens, and the third a resident alien, studying and teaching in the city for much of his life. The greatest Western historian of the period, Thucydides, was Athenian, and his successor, Xenophon, though an ardent admirer of Spartan militancy, was born and raised just over fifteen miles from Athens. Sophocles and Aristophanes, from whose minds flowed the most profound tragedy and biting satire in the literature of ancient Greece, were also natives of the city of Athena.

But what of the public at large? One particularly useful indication of the general level of learning in the city is the proportion of citizens who were literate. A variety of techniques have been used to estimate Athenian literacy, primarily centering on the reading required for participation in public life, the archeological evidence of writing on pottery fragments and the like, and references to reading in contemporary plays and prose works. By all accounts, Athens was the most literate society in the Western world at that time. William Harris, the most skeptical and influential recent writer on the subject, is at great pains to demonstrate that literacy was not as widespread in ancient times as had been previously thought, but even he relents somewhat in his discussion of Athens. He writes that "among the well to do, practically all males must have been literate" (Harris, 1989, p. 103). Harris neglects to offer an estimate of literacy among urban Athenian citizens, saying only that at least 15% of the male population as a whole, including the surrounding areas, was literate. Using his own data and arguments, it is fair to say that perhaps twice that percentage of city-dwellers were able to read, and most of these would have been able to write as well. Conversely, literacy among the rural population was probably at about half the overall level. This difference was due in large part to the greater frequency with which farming families required the labor of their children, thus leaving them fewer years during which to attend school. Similar constraints affected the urban poor, who had to apprentice their children to a craft at perhaps the age of 11 or 12.

Pedagogical freedom and market pressures both allowed and encouraged Athenian educators to make great strides. Independent Athenian schools were the first to introduce games as a pedagogical tool, and to reduce the use of corporal punishment-ubiquitous in Egypt and Sparta-to the exception rather than the rule. Elementary schools altered their curricula to meet changing parental demands, and an entirely new educational institution, secondary schooling, was brought into being as a result of market forces. In the words of Adam Smith:

> The demand for such [higher] instruction produced, what it always produces, the talent for giving it; and the emulation which an unrestrained competition never fails to excite, appears to have brought that talent to a very high degree of perfection. (Smith, 1994, p. 837)

These achievements, so far ahead of contemporary practice, went hand in hand with the spirit of freedom and community that pervaded Athenian society. Without resort to government intervention or coercion, Athens enjoyed not only an explosion of artistic, literary, and scientific work, but also a thriving economy. The depth and breadth of Athenian commercial life was by far the greatest of any city in Europe at the time, comparing favorably even with cities that existed centuries later. By allowing youths and adults to pursue a wide range of studies, the Athenians fostered a labor-market of exceptional diversity. The existence of skilled apprenticeships ensured a talented pool of craftsmen, while training in writing and mathematics made possible ever larger and more complex business transactions. Isocrates observed that "the articles which it is difficult to get, one here, one there, from the rest of the world, all these it is easy to buy in Athens." (Durant, 1939) In support of its vigorous shipping industry, Athens even offered a variety of financial and insurance services, which required both literacy and numeracy. As economic historian Rondo Cameron points out:

> Some cities, such as Athens, concentrated a number of commercial and financial functions within their boundaries in much the same way as Antwerp, Amsterdam, London, and New York did in subsequent eras. Banking, insurance, joint-stock ventures, and a number of other economic institutions that are associated with later epochs already existed in embryonic form in classical Greece (1993, p. 35).

The picture which comes down to us of Sparta in the 5th and 4th centuries B.C. is a very different one. Parents had no direct say in the education or upbringing of their children, having to cede their responsibilities and desires to a single, monolithic system. Innovations in language instruction and even physical training were suppressed by central control, leaving teachers without autonomy or flexibility. Sparta had virtually no science or literature, and little art. Her legacy to modern times is negligible, apart from being a beacon to totalitarians at the time of the French revolution and the rise of the Third Reich in Germany. Social stability, the result of voluntary association in Athens, was maintained by innumerable forms of government coercion and regulation, particularly in education.

Though one or two historians have attempted to show the existence of literacy among the common people in Sparta, there is a dearth of evidence to support their claims. Apart from the kings and perhaps a few generals and magistrates-who communicated with one another on "code sticks"-the Spartans were an illiterate people. Their economy was basic, and far more dependent upon slave and serf labor than that of Athens. The citizen class was allowed only to train for war in the state schools, and could neither acquire a broader learning nor apprentice themselves to skilled tradesmen. Trade was in fact actively discouraged by the Spartan government, in an effort to keep its people focused on an ascetic military lifestyle. In this, they were eminently successful.

### Germany and The Reformation

In a bustling German town, in the year 1500, a public notice proclaimed that "Everybody now wants to read and to write" (Schwickerath, 1904). Though this was still something of an exaggeration, it captured the spirit of the time. With the invention of the printing press, books became cheaper and more widespread throughout Europe, making literacy in the common languages of its people a practical and valuable skill for the first time in a thousand years. It also came within reach of a larger segment of the population, thanks to the diversification of the economy and the appearance of a small but growing middle class who could afford both books and teachers' fees.

Since the fall of the Roman Empire, education in the West had been the prerogative of the Catholic clergy, and Latin had been their language of choice. Naturally, as the demand for literacy grew, the middle classes turned first to this traditional seat of learning for instruction. Two factors soon changed this practice. The most notable was that an increasing number of citizens wished to learn German rather than Latin, and the church had little inclination to oblige them. As a result, the demand for German literacy was met by entirely private schools that introduced both children and adults to the perennial basics for a small fee. These popular independent schools spread rapidly in the larger towns, but were less numerous in villages and rural areas. The second cause of change in the provision of education was the desire of the public for greater control over the schools. As townspeople still favoring an education in Latin contributed more generously to their local parish educational funds, building new schools and retaining more teachers, they sought proportionately greater control over school staffing and curriculum. This did not sit at all well with the clerics who had until then been responsible for such decisions, and they often resisted any circumscription of their authority. Many considered it the fundamental right of the Church to control education. In the majority of cases, however, the citizens eventually won out, and city councils became the primary authorities over the schools formerly run by the clergy. Because clerics made up the vast majority of those capable of giving Latin instruction, most teachers in "city schools," as they came to be called, continued to be members of the clergy. School costs at these quasi-public institutions were paid for with a combination of tuition fees and taxes, broadening access, while still leaving some incentive for the students or their parents to ensure that they were receiving value for their money. The new trends towards private schooling and local community control were derailed, however, by one of the largest social upheavals in European history.

The Reformation threw German schooling into chaos. Schools staffed or run by the clergy closed down as monks and nuns abandoned their convental lives in droves. The process was accelerated by the nobility, who seized the opportunity to close all the monasteries that remained, excepting those that had adopted Protestantism. Finally, after several decades, new schools started to appear. Free enterprise elementary schools, which had been the least affected by the turmoil, were the first to recover. The printing industry had been central to the success of Protestant reform, and the demand for instruction in reading and writing that it had helped to spread remained strong. The efforts of private citizens to educate themselves were once again cut short, however, by one of Luther's close associates; a scholar named Melanchthon. Apparently believing that he knew what was best for the people, Melanchthon called for the creation of a government-run school system. With the help of Luther and the nobility of various German states, he was successful, and soon the existing private elementary schools were joined by state institutions. Because they were paid for by taxes rather than tuition fees, the new schools tended to make private instruction financially burdensome. Parents who wished to send their children to a private school had to pay both for it and for the state schools as well. Private schools were further discouraged by the attitudes and actions of the new state educational authorities, who derided and persecuted them (Paulsen, 1908). Attempts were even made to legislate private instruction out of existence (Cole, 1960), and in response they were sometimes forced to carry on their classes clandestinely. Though these "hedge schools" survived into the 17th and 18th centuries; they were marginalized by the growing state educational system.

Melanchthon's vision for mass education was inspired by the guiding principle of the reformation: the direct interpretation of the bible by individuals. The practice, however, was substantially different from its inspiration. If scriptural analysis was left to laymen, so the argument went, "incorrect" interpretations might result. The definition of what was incorrect was of course established by the leaders of the Reformation. As a result, reading, writing, and religion were taught using a pair of elementary catechisms composed by Luther. While he genuinely wished to improve the lot of children, Luther's views on what sort of education was acceptable were narrow and authoritarian. He felt that secular schools would lead to moral bankruptcy, and believed that parents should be compelled to teach their children according to his own views. Despite the spread of independent schools, he wrote to the reigning political authorities that: "It is to you, my lords, to take this task [education] in hand, for if we leave it to the parents, we will die a hundred times over before the thing would be done." (Chartier, 1976) Education once more became religious indoctrination, only this time it was legally mandated by the state. Fortunately for the majority of students who would not go on to a life in the clergy or government service, elementary instruction was given in their mother tongue.

The fate of Germany's city-schools was much the same as that of its private elementary schools. Political authorities at the state level were only slightly less hostile to local government institutions than they were to private enterprises. Pushed and squeezed by the state bureaucrats, city-schools found their curricula and attendance ever more limited. At the same time, new state- run institutions were created and given special privileges which the city-schools were not permitted to offer, such as the right to send their graduates on to university or into particular professions. Occasionally, city-schools were simply taken over by the state out of hand. In the late 16th and early 17th centuries, their pupils were mostly hand-picked by local lords, with the remaining openings allotted to the children of townspeople. Turning away from the popular movement towards education in German, and back to the classical languages so dear to the hearts of reformers, school regulations typically ordained that the new state secondary schools would teach in Latin. Their curriculum, too, culminated in the study of classical literature and scripture. Graduates were expected to converse fluently in Latin and have a passing acquaintance with Greek. In this end they were quite successful, but their achievement came at a cost to German culture and society.

Just prior to the Reformation there had been significant overlap in the education of the nobility and the training of at least the more avid youngsters from the middle classes. Education had been in the mother tongue for all but the clergy, and literate families in the towns and villages could and did share in the prose of their countrymen. Legal proceedings had also been held in German, allowing citizens to participate directly in any court actions which affected them. Once the strictly Latin secondary school system of the reformers was imposed, however, German gradually disappeared as a language of law and culture (Paulsen, 1908) This caused an ever greater rift between the uneducated masses and the learned elite which persisted for hundreds of years.

### On the Eve of the Modern World: England

After the civil wars of the mid 17th century, England was a country without a King. To cement their victory, the Puritan rebels abolished the House of Lords, withdrew the political powers of the bishops, and executed King Charles I on the grounds that his continued existence might encourage royalist revolt. They had little time to enjoy their new- found authority, however; as they were themselves deposed only eleven years later. In 1660 the monarchy was restored, and all its political and religious trappings with it. To forestall any further Puritan uprisings a host of restrictive laws were put in place against them. The Corporation Act of 1661 restricted public office to Anglicans, and it was quickly followed by the broader Act of Uniformity. Under this new legislation, educators at all levels were forced to sign a declaration of conformity to the Church of England's liturgy, and to give their oaths of allegiance to the crown. Nonconformists were thus prohibited from teaching in public and private schools, and their ministers were forbidden from coming within five miles of where they had once preached.

As political winds shifted over the next hundred years, the repressive religious and educational laws were at times ignored and at others reasserted. Having been forced to retreat from public life, the Puritans focused their energies on trade and commerce, expanding the middle class and thus the market for innovative schools. To satisfy this growing demand, a few private, fee-charging academies began to appear, founded illegally in many instances by non- conformist ministers who had been ejected from the teaching profession. In an effort to attract both dissenting and Anglican families, these schools offered an updated, predominantly secular curriculum with an emphasis on English, mathematics, and the natural sciences. One such school, operating in Tottenham in the 1670s, taught "geometry, arithmetic, astronomy and geography, with gardening, dancing, singing and music" in addition to English and some Latin (Lawson & Silver, 1973). Traditional endowed grammar schools, on the other hand, assured of a steady income independent of their ability to attract students, continued to provide the same classical Latin training they had offered since the Middle Ages. The polarization of these two forms of schooling, and their respective fates, clearly illustrate the role of market incentives in the educational process.

The continued growth and diversification of the economy dramatically widened the disparity between the content of traditional education and the needs of the commercial and professional classes. Together with the decline of the Church as an employer, this shift diminished whatever economic advantage the old syllabus might have conferred. Critics denounced the grammar schools as moribund and irrelevant, while parents increasingly sought more practical alternatives. As a result, the conservative endowed schools began to lose middle class pupils to the few private academies that had sprung up in the late sixteen-hundreds. Within a few decades this burgeoning change had solidified into a steep recession for traditional education, and a proliferation of new private academies. In the 18th century, grammar schools continued their descent, as few new ones were opened, some closed, and the rest saw their enrollments drop significantly. When Nicholas Carlisle conducted his multi-year investigation of hundreds of endowed schools in the early 19th century, he found many of them had lost touch with their prospective customers, and showed visible signs of decay. In Stourbridge, for example, he found that the school had taught only a trifling number of students over the preceding forty years, "as Classical learning is in little estimation in a commercial town." (Carlisle, 1818, v. II, p. 773) Despite the fact that Stourbridge's grammar school sometimes had no pupils at all, both its head and assistant masters continued to draw their full salaries. This was in fact not unusual, as masters, once awarded tenure and assigned a fixed salary, were virtually impossible to remove, even in cases of serious neglect (Lawson & Silver, 1973).

Endowed grammar schools were not entirely beyond the reach of market forces, however. In the many cases where the endowment was low, schoolmasters generally took the financially expedient steps of recruiting private pupils or taking on outside employment to increase their income, necessarily reducing the time they had for their endowment students. Others, such as those at Donington and Cuckfield, taught only one or two "free" (endowment) students, while conducting private lessons with scores of paying students on the foundations' premises (Carlisle, 1818, p. 345, 597). Finally there were masters who simply converted the school buildings into private residences, took no pupils of any kind, and continued to draw their stipend. Despite these systemic problems, there were schools led by dedicated masters able to make do with their allotted salary, that continued to instruct their pupils on the language and literature of ancient Greece and Rome. To the extent that endowed schools modernized their curricula to attract students, however, it was due primarily to the financial imperative.

In direct proportion to the decline in health and popularity of endowed grammar schools, private institutions grew and flourished. Subjects long ignored by the grammar schools began to appear, and soon entirely new ones were added. Arithmetic and geography were among the first, and these were joined by anatomy, biology, bookkeeping, economics, surveying, naval studies, and many others. While sometimes maintaining vestiges of the traditional curriculum, private institutions usually allotted them less time and importance than the new subjects. At St. Domingo House School, for example, Latin instruction was given but only after the children had received several years of training in French and German (Roach, 1986, p. 127). Not only were the subjects new, but the methods were often innovative as well. In keeping with the applied scientific nature of many of the courses, experiments using telescopes, microscopes and other devices complemented the familiar teaching methods. The teachers of Hill Top School conducted lessons with marbles to give children an intuitive grasp of arithmetic before introducing them to numbers and word problems. Physical surveying was used to teach trigonometry at the same institution (Roach, 1986, p. 124). One of the most concrete signs of the different attitude of the private schools was that many catered to girls, while grammar schools did not. Though the curriculum for girls was sometimes less academically ambitious, and always included ample emphasis on morals, manners, and domestic skills, it was at least a step forward.

For the very poorest families, who usually had no interest in a classical education and who could not afford the tuition at the better private institutions, two options remained; religious charity schools and private Dame schools. Though charity schools generally taught basic reading skills, they suffered from the same conflict of goals as the grammar schools. Just as the wealthy donors who endowed grammar schools generally insisted on a traditional Latin curriculum, the middle-class religious societies that funded charity schools had ideas all their own as to what the poor should learn, and these only rarely took into account the interests of the poor themselves. The central purpose was always to inculcate the moral and religious views of the sponsors. A widely held view among religious societies was that "Reading will help to mend people's morals, but writing is not necessary." (Smith, 1931, p. 53) An additional problem with religious charity schools was that the teachers were appointed and supported by religious authorities, rather than by the educational marketplace. Since those overseeing charity schools rarely had children attending them, there was little incentive for them to ensure the teacher's competency. Sometimes sound selections were nonetheless made, but in the worst cases masters were appointed who would never have been able to draw paying students. In Yorkshire, for instance, a "very deaf and ignorant" teacher was appointed by the parochial authorities "that he may not be burdensome to them for his support." (Lawson & Silver, 1973) Not surprisingly, the appeal of these schools was limited. Despite the fact that private schools charged tuition, "the subsidized, endowed and charity schools of Manchester attracted only 8 percent of all those attending schools and there were empty places available." (Royle, 1990)

The ubiquitous Dame schools, usually located in the home of an elderly local widow, also varied widely in quality based on the knowledge and skills of individual teachers. Competition generally kept the fees for such schools at a minimal level, however, and the freedom of families to chose among different teachers ensured that those who failed to meet their client's expectations could remain in business for only a short time. Despite their many shortcomings, Dame schools taught far more students from even the poorest classes than did charity schools, and, as we shall see below, they succeeded in most cases at conveying the rudiments the English language.

The major religious denominations were not entirely beyond the reach of competitive incentives, however, as is evidenced by the rise of the monitorial system. Monitorial schools, in which the brightest students taught all the rest, drew enormous interest around the turn of the 19th century due to their ability to reach far greater numbers of children at a lesser cost. A single schoolmaster, after imparting the day's lessons to his core of "monitors", could simply sit back and supervise as they carried out the bulk of the instruction. Of course, the quality of instruction depended on the presence of sufficient numbers of bright and capable students, and in some cases was probably only a small improvement over no education at all. Financially, however, the case was clear. The economy of having only one teacher for an entire school meant that formal education could reach even the poorest families. This ability to reach a much larger audience quickly caught the attention of the Church of England, in large part because the first monitorial schools had been run by a Quaker, Joseph Lancaster, along nondenominational lines. The prospect of having so many children educated in what was a predominantly secular environment was anathema to the Church, and so it set about creating its own monitorial system with the elephantine title of "The National Society for Promoting the Education of the Poor in the Principles of the Established Church." Wherever Lancaster had founded a school, the National Society created one of its own with which to compete. Soon the Church of England's network had grown vastly larger than that of its adversary. In keeping with its other educational efforts, the Church's monitorial schools were "instituted principally for Educating the Poor in the Doctrine and Discipline of the Established Church." (National Society, 1972, p. 50) These schools were not intended to provide children a stepping stone to higher studies, but rather to fit them to their positions at the bottom of the social and economic hierarchy. In strictly regimented lessons the pupils were taught to be satisfied with their subservient role in life. Due to this doctrinaire style and the curricular limitations imposed by the Church, monitorial schools failed to transform English education. Dame schools and other private ventures continued to reach a greater number of children than the religious charitable institutions (Royle, 1990).

By the second half of the 19th century, the governmental role in education had increased substantially. The main religious educational societies were now subsidized by parliament in an effort to improve the opportunities of the poor, and state inspectors visited their schools. Friction was high between Church and state over the proper distribution of regulatory and funding powers, and many within the government felt there was insufficient emphasis in the schools on basic subjects and younger grades. In 1862 a "Revised Code" for education was passed into law with the well-intentioned goal of bringing competition and the profit motive into education. The "Payment by Results" program, as it came to be known, stipulated that schools should be paid based on a combination of attendance and student performance on tests administered by state inspectors. What the Council members failed to understand was that by placing the financial strings in the hands of state inspectors instead of families, they would pull the attention of teachers and administrators away from the pupils and towards the government. Failing to satisfy the inspector meant a significant loss in funding, perhaps even forcing the school out of business, while receiving a positive review increased the institution's income. Student learning, insofar as it was not measured by the inspector, was of little financial consequence. The results were tragic.

Even before the legislation was passed a few observers warned that payment based on a few simple tests would encourage teachers to curtail their instruction in other subjects. In the event, these fears were fully realized. Years after the system had been put into practice, T. H. Huxley observed: "the Revised Code did not compel any schoolmaster to leave off teaching anything; but, by the very simple process of refusing to pay for many kinds of teaching, it has practically put an end to them" (Lawson & Silver, 1973). The testing system consisted of six separate levels, and since children could not be tested at the same level twice, or at a lower level from any previous attempt, schools held back older students so that they could be made to progress through all six levels, bringing in the maximum amount of cash over their educational lifetime. To ensure top scores at inspection time, teachers adopted frequent testing and memorization sessions. Often the children were made to learn their entire reading texts by rote so that they would have the least chance of failing. While some inspectors attempted to subvert these ploys by supplying an altered text or by asking the student to read backwards, others simply passed them: "I consider it to be my duty according to the letter of the Code, to pass every child who can read correctly and with tolerable fluency, whether he or she understand or not a single sentence or a single word of the lesson" (Smith, 1931). Reports from inspectors repeated the same criticism time and again, namely, that students were simply being made to memorize words without understanding their meaning. After years of experience with the system, the Cross Commission confirmed these views, faulting the teaching of reading under the Revised Code for being "too mechanical and unintelligent" (Vincent, 1989). Matthew Arnold (1972), the best known of the inspectors, summed up the consensus among his colleagues:

> I find in [English schools], in general, if I compare them with their former selves, a

> deadness, a slackness, and a discouragement... If I compare them with the schools of the continent I find in them a lack of intelligent life much more striking now than it was when I returned from the continent in 1859.

Not only the education but even the welfare of many children was sacrificed under this system. If a child was absent on the day of the inspection, even if gravely ill, the school would lose his or her attendance allocation. As a result it was not unheard of for school masters to compel children stricken with serious, even infectious, diseases to attend. One inspector observed that:

> To hear paroxysms of whooping-cough, to observe the pustules of small-pox, to see infants carefully wrapped up and held in their mothers' arms, or seated on a stool by the fire because too ill to take their proper places, are events not so rare in an inspector's experience as they ought to be. The risk of the infant's life, and the danger of infection to others, are preferred to the forfeiture of a grant of 6s. 6d. (Smith, 1931)

Teachers, forced by financial necessity to provide only the narrowest education to their students, lost all spirit and enthusiasm for their work. Their vocation had been reduced to a game of cat and mouse between the school and the inspector, in which teachers had to learn how to manipulate the system in order to be successful.

Despite its significant impact on schooling, the Revised Code was not the government's most lasting intervention into education. In 1870, W. E. Forster's Education Act added state provision of schooling to its existing roles in funding and inspection. Local school boards were created across the country to fill perceived gaps in the existing network of private and subsidized schools. Over the next several decades, state authority was progressively increased, attendance was made mandatory for children between ages 5 and 13, and tuition fees were gradually reduced to zero by 1918.

Analyzing the changes in literacy and student enrollment that occurred in the 19th century provides additional insight into the relative roles of independent and state schools. The most systematic evidence on literacy during this time period, both in England and elsewhere, is the frequency with which newlyweds signed their marriage documents-as opposed to simply making a mark. A strong argument can be made that this measure is more accurately described as a negative indicator of illiteracy, since the level of writing ability necessary for signing one's name is minimal, but its usefulness in the absence of other reliable statistical evidence is widely accepted. What these data show is that literacy increased steadily from 67.3% in 1841 to 93.6% in 1891, reaching 97.2% by 1900 (West, 1994). In interpreting this evidence it must be kept in mind that the difference between the mean school leaving age and the mean age of marriage was approximately 17 years. In other words, the 67.3% literacy rate already existing in 1841 cannot be attributed in any way to the initiation of state subsidization, which took place only 8 years earlier. Furthermore, the achievement of 94% literacy in 1891 was accomplished almost entirely before the Forster Education Act of 1870 had had time to generate an effect on the adult population. West has also shown that literacy was on the rise well before 1841.

The trend in school enrollment was substantially similar to that in literacy. The number of children in schools rose "from 478,000 in 1818 to 1,294,000 in 1834 `without any interposition of the government or public authorities.'" (West, 1994, p. 172) Between 1841 and 1850, the number of unsubsidized private schools grew from 688 to 3,754, while subsidized and endowed schools only increased from 415 to 616. Given the rapid rise in enrollment already under way prior to 1870, and the fact that subsidized Board Schools drew many of their customers away from existing private schools, West observes that it is difficult to discern any additional growth in enrollment that could be reasonably attributed to the Forster Education Act.

These figures, particularly for the early years of the 19th century, bear witness to the willingness of even the poorer and less well-educated parents to see to the education of their children, without state compulsion or supervision. Not only were poor parents sufficiently responsible to send their children to school, they also demonstrated a commendable level of selectivity among their various options. The relative failure of subsidized charity schools to attract parents, as compared to Dame and other fee-charging schools, indicates that parents were not only able to choose, but were willing to incur a financial burden in order to do so.

The behavior of teachers in private and subsidized schools is also telling. For more than a hundred years, the private academies of England were the only option for parents seeking a modern curriculum in language, technology, and science. The demand for practical instruction in accounting, surveying, applied sciences, naval skills, and other disciplines key to economic diversification and a higher standard of living were met almost entirely by private teachers. Tenured grammar school masters hung onto their limited Latin and Greek curriculum well beyond its period of usefulness, while religious charity schools often down-played the teaching of writing. Under the Revised Code, the incentive for subsidized-school teachers to satisfy the needs of families was further reduced, while a powerful new incentive to satisfy the baseline requirements of the inspectors was created, with dire results.

## France After the Revolution

French education, even more so than that of other European nations, was the battle ground for an epic religious and political power struggle. From monarchy to republic and back again, the revolutionaries strove to use the schools to shore up their position, vying for control with the firmly entrenched Catholic Church. It seems natural to suppose that on the eve of the revolution, with its emphasis on human rights and freedoms, the manipulation of education for political and religious ends would have lessened substantially. This, however, was not the case. The government that eventually emerged, while revolutionary in many respects, continued the age old tradition of using schools as a tool. In order to undermine the power of its primary opponent, the Catholic clergy, parliament severed all ties between education and religion. Nuns and priests were ordered to sign a constitution restricting their freedom to teach according to their faith. Since compliance with this order was difficult to achieve, the government soon resorted to a more direct approach: outlawing the clergy entirely. In one of history's more remarkable contradictions, the revolutionaries argued that a truly free nation could suffer no religious or secular societies amongst its citizens, and so abolished them (Chevallier, 1969). Simply wearing religious garb became a crime (Gontard, 1959).

Without a well-organized transitional strategy, schooling quickly began to collapse. Like Emperor Nero fiddling as Rome burned, the French parliament continued to debate exactly what the new system should look like as the old one crumbled around them. A genuinely revolutionary minority defended the right of families to choose their schools, whether sectarian or otherwise, but their voices were lost amidst a majority who believed the only choice was between moderate and absolute state control over education. So fervent was the belief in the power of the state and of the value of forced equality, that proposals for a totalitarian system much like Sparta's were put forward, in which children were to be taken away from their parents and educated in government communes. According to the delegate Le Pelletier, "The totality of the child's existence belongs to us [the state]; the clay, if I may express myself thus, never leaves the mold." (Ponteil, 1966)

Eventually a school law was passed, making attendance mandatory and requiring instructors to sign a "civic certificate" restricting their right to provide sectarian religious instruction. In place of the old catholic teachings, a new "natural religion" was imposed on the youth of France. Students were issued catechisms which admonished them to "worship Reason and the Supreme Being," in the deistic republican fashion (Barnard 1969). Having stripped away the traditional religious aspects of schooling, parliament had made teaching decidedly unattractive to the priests and nuns who comprised the vast majority of educators. The supply of willing teachers was thus reduced to a trickle. Even where teachers were to be found, many families resented both the intrusion of the state into their lives, and the ouster of Catholicism, and so kept their children at home. Though government policy had interrupted the existing supply of education, demand remained largely undiminished. So, in the gap created by the failure of state schools, independent religious institutions began to reappear. Unsurprisingly, these new schools were viewed by the republican parliamentary majority as strongholds of fanatics and royalists, to be "struck down" and "annihilated." The continued affinity of many citizens for traditional institutions was itself viewed as a sign of ignorance and lack of learning.

Ten years after the revolution the French educational scene looked like precisely what it was; a battle field. The general consensus of local officials and national observers was that an already weak system had been made worse. Report after report flowed into Paris, each lamenting the sad condition or complete absence of elementary schools. In the midst of this bleak educational landscape, a small group of philanthropists perceived what they thought might be an oasis. Having encountered and been impressed by English monitorial schools on a number of occasions, these men believed the system could help to circumvent the teacher shortage from which their country was suffering, while also replacing the outdated individual instructional technique with more effective group teaching. So, in June of 1815, the first French monitorial school was opened in Paris.

From its original handful of students the new school rapidly grew to an enrollment in the hundreds. Its success was widely praised and by the fall several other monitorial schools had appeared. Beyond the cost-effectiveness of the method, several of its pedagogical innovations attracted significant attention. Monitorial schools cast aside the existing practice of teaching reading and writing as entirely distinct skills, with excellent results. They furthermore grouped students by aptitude in each particular subject rather than strictly by age, allowing the children to progress through the curriculum at their own pace. Finally, in what seems an obvious move to modern readers, they taught to entire groups of students at once, rather than individually to each child in succession. The one-on- one method, wherein most of the class would devolve into chaos as the teacher focused his or her attention on a single student, had persisted in most church and state schools until the advent of the monitorial system. Of course critics aptly pointed out that the system tended towards excessive regimentation, but the problem was at least less severe than in the monitorial schools of England's National Society. In practice the advantages of the approach seem to have outweighed its weaknesses, for mutual instruction, as it became known, soon spread through France. By January of 1819 there were already 602 monitorial schools. Later that same year the number had increased an astounding 50%, to 912, and continued growing at that rate, reaching 1300 schools by February of 1820 (Gontard, 1959). Not only did the system succeed in opening more schools faster than any previous approach, it was in such great demand that many existing schools were forced to adopt its techniques in order to compete. "Instructors following the old method, seeing their pupils desert in order to run to the new one, are hurrying to adopt it themselves," observed a speaker at the general assembly in Paris (Gontard, 1959).

Unprecedented in their popularity with the citizenry, monitorial schools were nonetheless resented by the state and loathed by church. Managed and funded as they were by either secular private charities or municipal authorities, they enjoyed a significant measure of independence, making them difficult to manipulate by the established powers. The two most invidious characteristics of the system, as seen by Church and state, were its secularism and its meritocratic nature. Supporters of mutual education lauded the fact that it taught children "to obey merit... no matter who its repository may be," (Fouret & Ozouf, 1982) i.e. to disregard notions of social class, but the clergy argued that this would subvert the social order (Moody, 1978). The assembly and the University of Paris also feared they were losing their hold on education, and so set out to regain it.

In the years after its founding, the University of Paris had seen its role in primary and secondary schooling marginalized, and its influence atrophy. With education legislation pending in the assembly, its governors saw an opportunity to reassert their authority. This task proved somewhat easier than might be expected due to the fact that most of the of those drafting the legislation were prominent members of the University, committed to its control over all schools. The church was still a powerful force, however, and its lobbying won several compromises in the final law. The legislative patchwork thus created had bits to suit everyone, except, perhaps, the people of France: The University won a monopoly for granting the newly required teacher certifications; the Catholic Church was appeased by the requirement for thousands of regional supervisory committees, which its priests would head; and municipalities, due to their limited political influence, ended up with a few places on the Church's committees.

Though nominally meant to ensure the competence of candidates, teacher certification was entirely divorced from instructional practice. The examiners, usually local college professors selected by the University, had little knowledge of a primary school environment they had neither experienced themselves nor perhaps even observed (Ponteil, 1966). Usually too easy and sometimes too difficult, the uneven certification process was of little help in improving the quality of instruction.

Far more damaging than the haphazard certification of teachers was the requirement for regional school committees. Though headed up by local priests, these committees officially reported to the University, putting the Church in a subservient role. The clergy chafed at this limitation of their authority, and fought it with every technique they could devise. In a vast number of cases they simply refused to convene meetings, preferring to assume personal control over their local schools and school-masters. In those cases when the members did meet, internal squabbles were the norm, with the Catholic traditionalists and liberal defenders of mutual education locked in unswerving opposition to one another. Thanks to their organization and influence, the priests usually emerged victorious, picking whichever instructor best suited their needs. It was common for pious and acquiescent school-masters to receive favorable treatment, being freed from any legal requirements which might disqualify them from teaching, while those educators with strong individual wills, or with more liberal views, were persecuted and criticized in the priests' reports.

Committee members drawn from the local community were generally of little help in improving the process. Virtually all were otherwise employed and were neither willing nor able to spend a significant amount of time on the unsalaried position. With neither the experience nor the incentive to spur them on, their motivation quickly ebbed. Even proponents of the original law admitted its failure. In addressing parliament (Archives parlementaires, 1879), one of its founders, Guizot, made the following pronouncement:

> There are 2,846 cantons [in France]... For many years we have expended considerable effort organizing cantonal committees, but we have managed to create only 1,031; moreover, these still exist only on paper, there are hardly 200 that have taken any real action.

The final nail in the coffin of independent schools was the resurgence of Catholic political power. In the early 1820's the Church won an important victory, having bishop Frayssinous appointed Grand Master of the University of Paris, and Minister of Ecclesiastical Affairs and Public Education. From this new position of influence the Church was able to push through legislation granting it wide- ranging powers over teachers and schools. Classes were made to begin and end with prayers, its catechism was to be learned in daily lessons, and teachers were made increasingly answerable to the local priest. Due to their generally secular nature, and the fact that their origins lay in English Protestantism, monitorial schools were singled out for the fiercest attack. Priests leveraged their pulpits, demonizing mutual-teaching and its supporters in sermon after sermon. After only a few years of this new regime, monitorial schools were all but extinguished: their numbers were reduced from 1500 in 1821, to 258 by 1827 (Ponteil, 1966).

For the rest of the nineteenth century, the battle for control of education waged on. Though primary schooling reached an ever larger segment of the population, its nature at any given time continued to be decided by the faction with the greatest political clout. The degree of politicization and centralization of French schooling was well captured by the attitude of Hippolyte Fortoul, Minister of Ecclesiastical Affairs and Public Education from 1851 to 1856. Drawing a watch from his pocket he boasted that "At this moment, all the students of the lycees [secondary schools] are explaining the same passage from Virgil." (Moody, 1978, p. 59) Under Fortoul, the hours, methods, and content of teaching were all codified. Teachers were forced to swear an oath of loyalty, support official candidates, and were even prohibited from growing beards or mustaches.

Though the more liberal regimes of the eighties and nineties sought to make state education accessible to the entire nation, they stopped short of letting citizens decide exactly what kind of education was appropriate. Jules Ferry, nominated minister of public instruction in 1880, believed that all French children had the right to an education, but that the awarding of degrees must remain the prerogative of the state. This tool, coupled with the government inspection of all schools, was necessary in his eyes to maintain national unity and a common morality, and to regulate access to public office. Two national teachers' colleges, founded in 1883, insured a new generation of educators free from the conservative royalist views of the clergy. (Ponteil, 1966)

The traditional view of French educational history describes the 19th century as a period in which increased state intervention led to the expansion of schooling and the wider dispersion of literacy and culture. Certainly it has been shown that both state schooling and literacy grew significantly during the 1800's. Grew and Harrigan go somewhat further, concluding that since the correlation between enrollment and later literacy is larger than the correlation between literacy and later enrollment, state schooling must have been responsible for some of the growth in literacy (1991, p. 72). Even this cautious conclusion is subject to question, however. While Grew and Harrigan based their conclusion on the literacy figure for a single year, a study conducted by Furet and Ozouf (1982) looked at the literacy data at several points during the 19th century. Among their findings was that literacy was widespread in many Northern and Eastern districts in the 1700s, well before the appearance of state elementary schools. They also found that in general, areas that had high levels of state school enrollment already had high levels of literacy before that enrollment could have had an effect. Enrollment of 8 to 12 year olds in 1850, for example, was already strongly correlated with adult literacy in 1854. In other words, high levels of literacy and state school enrollment tended to be contemporaneous. Furet and Ozouf concluded that the relationship between literacy and schooling was to a great extent circular; literate parents were more likely to seek education for their children, and educated children were more likely to become literate. The entire process stemmed from a growing demand on the part of the public for literacy, spawned by the spread of written material and the increasing economic value of reading and writing. They wrote that:

> In the long term, [schooling] is nothing but a product of the demand for education. Of course, a school founded purely out of individual generosity or at a bishop's initiative may produce a temporary improvement in education in a parish; but its chances of enduring and of generating far-reaching changes in cultural patterns are slim, unless it is not only accepted but actively wanted by the inhabitants. (p. 66)

The truth of this assessment is attested to by the success of the independent monitorial schools, which not only flourished in response to popular demand, but led existing institutions to emulate their innovations. In many cases, these innovations were subsequently discarded by the state schools. The practice of grouping students by ability, for instance, though supported by modern research (Kulik, 1992), is rarely seen in schools to this day.

The battles over control of French schooling did have a significant impact on social stability, however. In the very area in which many educators tout the superiority government schooling over competitive market provision-fostering understanding and social harmony-the outcome appears to have been quite the opposite. Whether by republican parliamentarians or Catholic monarchists, the state schools were used as a weapon with which to bludgeon their opponents. In their time in office, the revolutionaries cut the clergy's ties to education in order to weaken their influence on the people. As the Church rose once again to power, Catholic teachings were legally forced on the state schools and private secular institutions came under heated attack. In contrast to this state compulsion, the independent monitorial schools placed no religious restrictions on their pupils or teachers. They were also the first to integrate children of upper and lower classes, but far from being supported in this by the educational bureaucracies of clergy and government, they were fiercely opposed.

## Conclusion

Having described the history of schooling in these four different contexts, it is useful to see what commonalities present themselves. In particular, it is fruitful to look back at the three measures of quality listed in the introduction, namely: responsiveness and innovation, direct benefits, and indirect benefits.

There is no question that competitive educational markets have been more responsive to the needs and demands of parents than centrally controlled, subsidized systems. This has held true whether the monolithic systems have been run and paid for by governments, as was most commonly the case, or by religious societies. In Athens, changing public demand resulted in changes to the elementary curriculum, and even led to the creation of secondary education. Spartan schooling, both due to implicit features of its organization and to the explicit wishes of its rulers, kept all innovation and progress at bay for hundreds of years. In pre- reformation Germany, it was the small private school that was first to offer instruction in the vernacular, both to adults and children. The state-run schools fostered by Luther and Melanchthon often ignored the wishes of the public, insisting on a classical course of studies useless to the common man. The same was true of England's endowed grammar schools. English Dame schools, by contrast, taught only what parents were willing to pay for, even attracting families away from the subsidized schools run by religious societies. For centuries, the most sophisticated and modern instruction in England was to be had at private secondary schools, which introduced the sciences, practical engineering and surveying techniques, naval skills, and living foreign languages. Before they were squeezed out of existence by tax-subsidized public schooling, there was simply nothing that could compare to them. In France, monitorial schools led the way in pedagogical innovation and in meeting public demands--so much so that other schools were forced to adopt their methods in order to avoid losing pupils.

In looking at the direct benefits bestowed on students by different approaches to educational organization, the clearest distinction to be found is between the practical and the pointless. Privately financed and operated schools have tended to offer programs of practical benefit to their clients, while centralized systems have taught only those subjects chosen by their founders or administrators--in most cases subjects of little value to the average member of the public. While private schools have consistently taught literacy in the vernacular of their clients for thousands of years, this has only rarely been the case in state or charity-run schools. When it was finally taught by the religious societies in England, they often deliberately omitted teaching writing. Similarly, practical training in mathematics and science has been ignored by bureaucratic school systems until quite recently, while their history dates back to the 5th century B.C. in private schools.

Perhaps the most glaring contradiction between the beliefs of modern public school advocates and the historical evidence is in the area of indirect or social benefits (also called positive externalities). Defenders of public schooling argue that only it can preserve social harmony and a sound economy, while a competitive educational market would lead to social strife and presumably economic deterioration. Nothing could be further from the truth. Government-run schools have in fact been far more coercive, and far more likely to lead to social discord than their private counterparts. Tying themselves to a single religion or ideology, public schools have often alienated all those who did not share the chosen views. When French monitorial schools encouraged the intermingling of children of different social classes, and respecting intellectual merit no matter what its source, they were actually criticized for it by the ruling powers of public schooling. When English law forbade non-conformists to teach, they taught nonetheless, privately and illegally, and generally admitted students irrespective of their religion. Because private schools allowed families the option of pursuing the particular kind of education they value, conflicts were avoided.

Whenever the state chooses one world view over all others, it places its own people into conflict with one another. This has been happening for centuries, and it continues to happen today. As for indirect economic benefits, there is simply no question. By offering more practical preparation than their government-run counterparts, private schools have contributed far more, per capita, to bolstering their national economies.

One area in which both private and public schools have performed poorly throughout history, at least by modern standards, is the provision of education for the poor. While it is possible to trace an historical desire among wealthy individuals to contribute to the education of the poor, this desire has rarely been effectively translated into action. Government-subsidized schools, as well as private religious charities, provided easier access to educational services than unsubsidized private institutions, but these services were not generally based on the needs and demands of the families they served. This is evident from situations such as the one in Manchester where free and subsidized schools held only a small share of the market, and, despite having empty places available, still lost potential customers to their unsubsidized competition. To a certain extent, poor parents have thus had to choose between the private schools that met their needs and the subsidized schools they could more readily afford, with little intersection between the two.

The import of the historical evidence for modern schooling is clear. Competition and the profit motive must be reintroduced into education so that teachers and school administrators will once again have a powerful incentive to meet the needs of the children and parents they serve. It can also be expected that the elimination of existing educational monopolies will alleviate many of the ongoing battles over curriculum and religion in the schools, by allowing families to pursue an education in accordance with their own values, without the need to impose those values on others. What remains to be resolved is the question of how to integrate the reintroduction of market forces with the subsidization of families with limited financial means. Vouchers and tax-credits no doubt offer a viable approach to the problem, though the need for more work in the design and application of these plans is paramount.

## References

Addonizio, M. (1994) School Choice: Economic and Fiscal Perspectives. Policy Report PR-B12 (Indiana Education Policy Center, Unpublished: ERIC Document # 371 447)

Archives parlementaires (1879) Archives parlementaire de 1787 a 1860, ser. 2, v. 83 (Paris, Librairie administrative de Paul Dupont)

Arnold, M. (1972) Report of the Committee of Council on Education, 1867-8, in: Goldstrom, J. M. (Ed.) Education: Elementary Education 1780-1900 (New York, Barnes & Noble)

Barnard, H. C. (1969) Education and the French Revolution (Cambridge, The University Press)

Bell, D. (1987) The Case for a Separate Black School System, Urban League Review, v 11, n 1-2, pp. 136-145

Blum, V. C. (1985) Private Elementary Education in the Inner City, Phi Delta Kappan, v 66, n 9, p. 645

Boaz, D. (1991) The Public School Monopoly: America's Berlin Wall, in: Boaz, D. (Ed.) Liberating Schools (Washington, D.C.; The Cato Institute)

Borland, M.V. & Howsen, R.M. (1993) On the Determination of the Critical Level of Market Concentration in Education, Economics of Education Review, v 12, n 2, pp. 165-169

Brown, F. (1992) The Dutch Experience with School Choice: Implications for American Education, in: Cookson, P. W. (Ed.) The Choice Controversy (Newbury Park, Corwin Press)

Cameron, R. (1993) A concise economic history of the world: from Paleolithic times to the present (New York, Oxford University Press)

Carlisle, N. (1818) A Concise Description of the Endowed Grammar Schools in England and Wales (London; Baldwin, Cradock, and Joy)

Carnegie Foundation for the Advancement of Teaching (1992) School Choice (Princeton, Carnegie Foundation)

Chandler, T.A. (1986) Research Points the Way to Cost- Effective Instruction, American School Board Journal, v 173, n 12, p. 48

Chartier, R.; Julia, D.; Compere, M. (1976) L'Education en France du XVIe au XVIIIe Siecle (Paris, Societe D'Edition D'Enseignment Superieur)

Chavis, B. (1994) A Native American Perspective on Choice, in: Billingsley, K. L. Voices on Choice (San Francisco, Pacific Research Institute for Public Policy)

Chevallier, P.; Grosperrin, B.; et Maillet J. (1969) L'Enseignment Francais de la Revolution a nos jours (Paris, Editions Mouton)

Childs, S. & Shakeshaft, C. (1986) A Meta-Analysis of Research on the Relationship Between Educational Expenditures and Student Achievement, Journal of Education Finance, v 12, n 3, pp. 249-63

Cole, L. (1960) A History of Education: Socrates to Montessori (New York; Holt, Rinehart and Winston)

Coleman, J. (1990) Equality and Achievement in Education (Boulder, Westview Press)

Cookson, P. W. (1994) School Choice: The Struggle for the Soul of American Education (New Haven, Yale University Press)

Coons, J. E. (1991) Perestroika and the Private Provider, in: Boaz, D. (Ed.) Liberating Schools (Washington, D.C.; The Cato Institute)

Durant, W. (1939) The Life of Greece (New York, Simon & Schuster)

Durant, W. (1957) The Reformation (New York, Simon & Schuster)

Elmore, R. (1990) Choice as an Instrument of Public Policy, in: Clune, W. H. and Witte, J. F. (Eds.) Choice and Control in American Education (London, Falmer Press)

Fossey, R. (1994) Open Enrollment in Massachusetts: Why Families Choose, Educational Evaluation and Policy Analysis, v 16, n 3, pp. 320-334

Freeman, K. (1904) Schools of Hellas (New York, Teachers College Press)

Furet, F. & Ozouf, J. (1982) Reading and Writing: Literacy in France from Calvin to Jules Ferry (Cambridge, The University Press)

Gontard, M. (1959) L'enseignment Primaire en France, 1789- 1833 (Paris, Belles Lettres)

Grew, R. & Harrigan, P. (1991) School, State, and Society: The Growth of Elementary Schooling in Nineteenth-Century France-A Quantitative Analysis (University of Michigan Press, Ann Arbor)

Hamilton, E. (1957) The Echo of Greece (W. W. Norton & Company, New York)

Hanushek, E. A. (1986) The Economics of Schooling: Production and Efficiency in Public Schools, Journal of Economic Literature, v 24, pp. 1141-1177

Hanushek, E. A. (1989) The Impact of Differential Expenditures on School Performance, Educational Researcher, May, 1989, pp. 45-51

Harris, W. V. (1989) Ancient Literacy (Harvard University Press, Cambridge)

Hoffer, T.; Greeley, A. M.; & Coleman, J. S. (1990) Achievement Growth in Public and Catholic Schools, in: Coleman, J. S. Equality and Achievement in Education (Boulder, Westview Press)

Isocrates (1982) Panathenaicus, in: Isocrates in Three Volumes with an English Translation by

George Norlin (London, Wiliam Heinemann)

Kozol, J. (1992) Flaming Folly, The Executive Educator, v 14, n 6, pp. 14-19

Krashinsky, M. (1986) Why Educational Vouchers May Be Bad Economics, Teachers College Record, v 88, n 2, pp. 140-151

Kulik, J. (1992) An Analysis of the Research on Ability Grouping: Historical and Contemporary Perspectives. Storrs, CT: National Research Center on the Gifted and Talented, The University of Conneticut.

Lawson, J. & Silver, H. (1973) A Social History of Education in England (London, Butler & Tanner Ltd.)

Levin, H.M.; Glass, G.V.; & Meister, G.R. (1987) Cost- Effectiveness of Computer-Assisted Instruction. Evaluation Review, v 11, pp. 50-72

Levin, H. M. (1991) The Economics of Educational Choice, Economics of Education Review, v 10, n 2, pp. 137-158

Lieberman, M. (1989) Privatization and Educational Choice (New York, St. Martin's Press)

Lieberman, M. (1991) Profit-Seeking Schools, in: Boaz, D. (Ed.) Liberating Schools (Washington, D.C.; Cato Institute)

Marrou, H. I. (1965) Histoire de l'education dans l'antiquite (Paris, editions du seuil)

Martinez, V.; Thomas, K.; & Kemerer F. R. (1994) Who Chooses and Why: A Look at Five School Choice Plans, Phi Delta Kappan, v 75, n9, pp. 678-81

Murnane, R. J. (1983) The Uncertain Consequences of Tuition Tax Credits: An Analysis of Student Achievement and Economic Incentives, in: James, T. and Levin, H. M. (Eds.) Public Dollars for Private Schools (Philadelphia, Temple University Press)

National Society for Promoting the Education of the Poor in the Principles of the Established Church (1972) First Report, 1812, in: Goldstrom, J. M. (Ed.) Education: Elementary Education 1780-1900 (New York, Barnes & Noble)

Organization for Economic Cooperation and Development (1994) School: A Matter of Choice (Paris, OECD)

Pande, R. (1977) A Study of Factors Affecting Parental Selection of A Primary School for Their Children in Tansen Town (Lalitpur, Amita Panday)

Payne, R. (1993) Poverty Limits School Choice in Urban Settings, Urban Education, v 28, n 3, pp. 281-299

Plato (1937) Protagoras, in: The Dialogues of Plato, Translated Into English by B. Jowett (New York, Random House)

Pliny The Younger (1969) Letters, Book IV, XIII, in: Letters and Panegyricus [by] Pliny (London, Wiliam Heinemann)

Plutarch (1988) Lycurgus, in: Plutarch on Sparta, Translated by J. A. Talbert (London, Penguin)

Ponteil, F. (1966) Histoire de L'Enseignment en France , Les Grandes Etapes, 1789-1964

(Paris, Sirey)

Rhinehart, J. R. & Lee, J. F. Jr. (1991) American Education and the Dynamics of Choice (New York, Praeger)

Roach, J. (1986) A History of Secondary Education in England, 1800-1870 (New York, Longman)

Royle, E. (1990) Modern Britain, a Social History 1750-1985 (Kent, Edward Arnold)

Schmidt, P. (1995) Looking the Other Way, Education Week, v 14, n 21, pp. 23-27

Schwickerath, R.(1904) Jesuit Education (St. Louis, B. Herder)

Sextus Empiricus (1987) Against the Professors, Book II, in: Sextus Empiricus with an English Translation by the Rev. R. G. Bury (London, Wiliam Heinemann)

Shanker, A. & Rosenberg, B. (1992) Politics, Markets and American Schools: A Rejoinder, in: Kane, P. R. (Ed.) Independent Schools, Independent Thinkers (San Francisco, Jossey-Bass)

Smith, A. (1994) The Wealth of Nations (New York, The Modern Library)

Smith, F. (1931) A History of English Elementary Education 1760-1902 (London, University of London Press)

Smith, K. B. & Meier, K. J. (1995) The Case Against School Choice: Politics, Markets, and Fools (Armonk, M. E. Sharpe)

Sowell, T. (1993) Inside American Education (New York, The Free Press)

Suetonius (1992) Suetonius, With an English Translation by J. C. Rolfe (London, Harvard University Press)

Traill, H. D. & Mann, J. S. (Eds.) (1909) Social England (London, Cassell & Co.)

U.S. Department of Education (1993) Digest of Education Statistics 1993 (Washington D.C., U. S. Government Printing Office)

U.S. Department of Education (1995) Use of School Choice, Education Policy Issues: Statistical Perspectives, May 1995 (Washington D.C., U.S. Government Printing Office)

Vincent, D. (1989) Literacy and popular culture: England 1750-1914 (Cambridge, Cambridge University Press)

Wells, A. S. & Crain, R. L. (1992) Do Parents Choose School Quality or School Status? A Sociological Theory of Free Market Education, in: Cookson, P. W. (Ed.) The Choice Controversy (Newbury Park, Corwin Press)

West, E. G. (1994) Education and the State: a Study in Political Economy (Indianapolis: Liberty Fund)

Williams, P. (1994) School Choice Promotes Excellence in the African American Community, in: Billingsley, K. L. Voices on Choice (San Francisco, Pacific Research Institute for Public Policy)

Winston, C. (1993) Economic Deregulation: Days of Reckoning for Microeconomists, Journal of Economic Literature, v 31, p. 1263-1289.

Xenophon (1988) Spartan Society, in: Plutarch, Plutarch on Sparta, Translated by J. A. Talbert (London, Penguin Books)

## About the Author

**Andrew J. Coulson**

***a_coulson@msn.com***

Andrew J. Coulson is an independent scholar based in Seattle, WA. Determining how schools can best be structured in order to serve the needs of families and communities has been the focus of his work over the past three years. He has written articles on the organization, condition, and history of education. His most recent publication, "Schooling and Literacy Over Time: The Rising Cost of Stagnation and Decline," appeared in vol. 30, no. 3 (October 1996) of the journal *Research in the Teaching of English*.

At present, he is completing a manuscript for the general public on the organization of schooling, tentatively titled *On the Way to School*. This book will address the educational problems currently confronting parents by clearly explaining what people want from their schools and how they can get it. It will do this by comparing school systems throughout history and showing which have worked, which have not, and why.

Prior to entering the field of education several years ago, Andrew Coulson was a systems software engineer with Microsoft corp. So, while Bill Gates quit school to form Microsoft, Andrew Coulson quit Microsoft to reform schools. He received his B. Sc. Degree in Mathematics and Computer Science from McGill University in Montreal, Canada (Andrew, that is, not Bill).

---

**Copyright 1996 by the *Education Policy Analysis Archives* and Andrew J. Coulson**

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" at olam.ed.asu.edu.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@zodiac.rutgers.edu*

Andrew Coulson
*a_coulson@msn.com*

Sherman Dorn
*dornsj@ctrvax.vanderbilt.edu*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

William Hunter
*hunter@acs.ucalgary.ca*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Dewayne Matthews
*dm@wiche.edu*

Les McLean
*lmclean@oise.on.ca*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Richard C. Richardson
*richard.richardson@asu.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

John Covaleskie
*jcovales@nmu.edu*

Alan Davis
*adavis@castle.cudenver.edu*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Ernest R. House
*ernie.house@colorado.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 10** **June 30, 1996** **ISSN 1068-2341**

---

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# Being Popular About National Standards: A Review Of *National Standards in American Education: A Citizen's Guide.*

Diane Ravitch.*National Standards in American Education: A Citizen's Guide.* Washington: The Brookings Institution, 1995. pp. 223. $22.95 (hardcover)

**Michael W. Apple**

**University of Wisconsin, Madison**

*APPLEMW@macc.wisc.edu*

**Abstract:** I assume that Diane Ravitch is someone who is as deeply committed to a fair and socially just education as I am--even when our political and educational agendas may differ--I also assume that re-stratification and fostering the power of the conservative restoration is not what she wants either. Thus, I do urge you to read this book, but perhaps for different reasons: to see it as a cautionary tale and then to watch as the public policies that are justified under its rhetorical umbrella and that are actually implemented on the ground go in uncomfortable directions.

Before you read any further, you should know that this will not be a "disinterested" review by a "disinterested" observer. Diane Ravitch and I have a prior history of interaction in print. Thus when her book written with Chester Finn--*What Do Our 17-Year- Olds Know?* (1987)--appeared I was invited to review it for a major journal. While I thought that the volume did raise some interesting issues, I also argued that it was flawed and was ideally suited to advance the neo-conservative attack on schools. Diane Ravitch responded, partly in a serious way but also in a relatively "cute" way that did not deal with the substantive concerns I raised, perhaps because of the length limitations imposed on any response. Through it all, it was clear that we disagreed in truly major ways. But, even with these substantial disagreements, the discourse never became that form of character assassination that too often poses as arguments between left and right.

At the risk of seeming consistent, I have exactly the same reaction to Ravitch's recent volume on national standards as I did to her earlier book on testing. Once again, it raises some interesting issues and once again I believe that its arguments are deeply flawed. This volume too is ideally suited to support political and cultural positions that are more conservative than Ravitch herself may be.

*National Standards in American Education* is meant to be a popular book. I do not mean this at all negatively. Educational policy and practice have become ever more complicated and strikingly political. Thus, there is a great need for books that sort through the complexities, present clear syntheses of different positions, and clarify what is at stake when particular positions are taken. Yet, because of this, authors of popular books have a real political and ethical responsibility to their readers. They must clarify, yet not overly simplify. They must do justice to positions about which they have serious disagreements. The task of the popularizer is to make arguments accessible, without creating caricatures-- straw-persons--whose arguments are but pale reflections of their original depth and power. Therefore, writing popular books on important issues requires an immense amount of discipline, not only stylistically but in reading and presenting the substantive arguments for or against one's position on educational policy carefully.

These requirements make me more than a little nervous about what Diane Ravitch has done--and has not done--in this book. Ravitch is indeed a fine writer. Her style is clear and unmystified. She has a nice way with words. However, she is considerably less successful in the other demands placed upon the popular writer. She all too often doesn't deal with either the best or the most rigorous arguments of those who do not agree with her presuppositions, often preferring to deal with only the somewhat rhetorical and brief statements of opponent's positions. Whether this is conscious or not, this is quite a clever strategy. It enables the "naive" reader to think that the author is being fair and equitable, at the same time that some telling points made by opposing arguments can be all too easily dismissed. (This is not only a problem with those whose educational, ideological, and political positions are similar to those of Ravitch. Unfortunately, this strategy is also found among those whose positions are closer to my own.)

Given the intense conflict over educational policy now--when it is crucial to listen carefully to multiple arguments about who benefits from the ways our curricula, pedagogies, and evaluation mechanisms are organized and controlled--I worry about this in general. But, in the case of this book my worries are more specific, since Ravitch has done this to my own writing as well as that of others. For example, as some of you may know, I have written at length about the movement toward national curricula, national standards, and national testing. I have raised a number of questions about its overt and hidden effects, its social and cultural claims, and its position on a "common culture" (Apple, 1992; Apple, 1993b).

In general, I have argued--along with many others--that the results of this movement will be that it will be captured by neo- liberal and neo-conservative tendencies and used for purposes whose large scale effects will be damaging to those with the least economic, political, and cultural power in the United States. I have also argued that many of these kinds of proposals are based on little understanding of the daily lives of teachers and the already intensified conditions under which they work. In even more recent work (Apple, 1996), I have brought to bear powerful empirical evidence--much of which was available even when Ravitch was writing this book--to demonstrate these effects. Yet, the representation of my arguments is taken from a two page piece written for a popular political magazine, a piece that was simply meant to provide something of a beginning point to make the reader aware of a set of issues, not to fully argue about them.

Ravitch wrote National Standards while in residence at The Brookings Institution in Washington. As with many of these kinds of think tanks, it too has moved significantly to the right. Thus, the political center has been redefined, often to such an extent that what earlier would have been considered to be quite a conservative position has often now become "moderate." This signifies a major transformation in our commonsense. Much of our public discussion involves quite simplistic neo-conservative versions of the issue of a "common culture." Increasingly, at the same time, other elements that surround what has been called the "conservative restoration" are becoming dominant. Thus, public is seen as bad and private as good. More and more, the neo-liberal emphasis on the marketplace as the ultimate arbiter of justice has been taken as "truth." Indeed, our very idea of democracy is in the process of being transformed. The citizen is now replaced by the individual consumer (See Apple, 1993a; Apple, 1996). And our ethical sensibilities are withering so that many people have now become almost inured to the human suffering that is produced by the ways in which our institutions operate--a reality that may be best described by Jonathan Kozol's powerful phrase "savage inequalities" (Kozol, 1991). While many of us lament this fact, my basic point is to remind the reader that Ravitch's book was itself written under a particular political aegis. It needs to be situated within a set of larger movements, not as an isolated volume about one part of educational life.

Basically, Ravitch is strongly in favor of national standards. These are to remain voluntary and dynamic, not mandatory and static. They are to be assessed in multiple ways, with a focus on that latest buzz word, performance assessment, not multiple choice tests. These kinds of examinations should be given to all individual students in a way that provides comparative performance data on similar students of the same age and grade level. Accompanying this will be the creation of report cards for individual schools and districts. Such clarified national standards and more detailed performance assessments will help colleges and universities and will assist employers. Employers will rely on high school transcripts and there will be a closer connection between what schools focus on and the skills needed to "succeed in the workplace."

There are elements of insight here: the voluntaristic nature of any standards that may be developed; the reduced emphasis on simplistic paper and pencil standardized tests; the urge to give "the public" more information about what schools are doing; the need to communicate to students and parents that education is very important; and so on. Yet, for all of her evident insights, it is almost as if Ravitch lives in an unreal world at times. Among the most powerful driving forces in American education at this time are increasingly something that sounds suspiciously like Social Darwinism and an impulse to use schools for re-stratification. At the same time, neo-liberal, neo-conservative, and authoritarian populist religious fundamentalists have created a tense but effective alliance in which market plans are coupled with proposals for national curricula and national testing. In essence, by putting in place national standards and then national performance testing, we can then set the market loose, since "consumers" will then have sufficient information to be able to choose among "products" (or schools). As odd as it may seem at first glance, the centralizing and rationalizing impulses of national curricula and national testing may be essential first steps toward the long term goal of marketization and privatization of schools through choice and voucher plans (Apple, 1996). This combination of strong state/weak state is exactly what is being tried in a number of nations under the new conservative policies being implemented. As Whitty and others have shown, the results have been more than a little undemocratic or very contradictory (Whitty, Edwards, and Gewirtz, 1993; Whitty, in press; Pollard, et al., 1994). Why should we expect that the US will be any different?

Of equal importance, is the fact that the fiscal crisis now being experienced in many states has meant that seemingly fine sounding plans--sometimes quite similar to what Ravitch has asked for--have served as excuses to put in place much of what she is against. Thus, for example, in a number of states--even after a good deal of work was done on higher standards and on more flexible forms of assessment--money was only allocated by the state for standardized, reductive paper and pencil tests. It was too expensive to do otherwise. The rhetoric of higher standards and of more flexible modes of assessment coupled with the fear of "declining economies" and "declining achievement" created a sense of urgency to get more testing in schools. However, the rhetoric of "higher" and "flexible" ultimately functioned to increase the power of mandatory state-centered testing of a relatively reductive kind, at the same time as there continued to be no growth in the ability of schools to do anything more about even meeting the old standards and tests. It ultimately functioned to add one more way of intensifying teachers jobs and of blaming the school even more for the social dislocations of this society. Speaking as bluntly as I can, my own prediction is that one of the most powerful and damaging effects of the standards movement and of the performance assessment movement will be to affix labels on poor children that will be even harder to erase than before.

I could go on here. But my basic point is a simple one. Diane Ravitch is quite a good writer and is able to make what seems to be an articulate case for higher national standards and more emphasis on performance assessment of particular kinds. However, she does this by simplifying the contentious issues, by ignoring important counter-evidence, and by failing to fully understand some of the most powerful economic, ideological, and political currents in the United States and elsewhere.

*National Standards in American Education* could perform a valuable service if it was read as a set of arguments about what to be very cautious of not doing in our drive to "reform" education. There are valuable issues raised in it. However, I predict it will be put to exactly the opposite use. It will add support to those neo-conservatives who wish to centralize control over "official knowledge" or by neo-liberals who want to reindustrialize the school by making schools into places whose primary (only?) function is to meet the needs of the economy and who see students not as persons but only as future employees. And this will occur at the very same time as major corporations are shedding thousands upon thousands of workers, most of whom did quite well in school, thank you very much. It will be used once again to export the blame for our economic and social tragedies onto schools, without providing sufficient support to do anything serious about these tragedies. And, finally, it will be used to justify curricula, pedagogic relations, and mechanism of evaluation that will be even less lively and more alienating than those that are in place now. (For alternatives to these kinds of things and to those that are proposed by Ravitch, see Ladson- Billings (1994) and Apple and Beane (1995)).

Do not misconstrue what I am saying here. As I have argued elsewhere, I am not in principle opposed to national standards or to the processes of assessment--if and only if they are employed to instigate a national debate at every school and in every community about what and whose knowledge should be considered "legitimate" and about the very real patterns of differential benefits our schools produce (Apple, 1996). If they do not do this, then they should be approached critically and with immense caution. Since I assume that Diane Ravitch is someone who is as deeply committed to a fair and socially just education as I am--even when our political and educational agendas may differ--I also assume that re-stratification and fostering the power of the conservative restoration is not what she wants either. Thus, I do urge you to read this book, but perhaps for different reasons: to see it as a cautionary tale and then to watch as the public policies that are justified under its rhetorical umbrella and that are actually implemented on the ground go in uncomfortable directions.

## References

Apple, M. W. (1992) Do the standards go far enough? *Journal of Research in Mathematics Education* 23 (5), pp. 413-431.

Apple, M. W. (1993a) *Official knowledge.* New York: Routledge.

Apple, M. W. (1993b) The politics of official knowledge: Does a national curriculum make sense? *Teachers College Record* 95 (2), pp. 222-241.

Apple, M. W. (1996) *Cultural politics and education.* New York: Teachers College Press.

Apple, M. W. and Beane, J. A. (Eds.) (1995) *Democratic schools.* Washington, DC: Association for Supervision and Curriculum and Instruction.

Kozol, J. (1991) *Savage inequalities.* New York: Crown.

Ladson-Billings, G. (1994) *The dreamkeepers.* San Francisco: Jossey-Bass.

Pollard, A. et al. (1994) *Changing english primary schools?* London: Cassell.

Whitty, G. (in press) Creating quasi-markets in education. In Apple, M. W. (Ed.) *Review of research in education.* volume 22. Washington, DC: American Educational Research Association.

Whitty, G., Edwards, T. and Gewirtz, S. (1993) *Specialization and choice in urban schools.* New York: Routledge.

**About the Author**

Michael W. Apple is John Bascom Professor of Curriculum and Instruction and Educational

Policy Studies at the University of Wisconsin, Madison. Correspondence may be sent to Professor Michael W. Apple, University of Wisconsin, Madison, Department of Curriculum and Instruction, 225 North Mills Street, Madison WI 53706 or via email at APPLEMW@macc.wisc.edu

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@zodiac.rutgers.edu*
John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Sherman Dorn
*dornsj@ctrvax.vanderbilt.edu*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 11** **July 24, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# National Education 'Goals 2000': Some Disastrous Unintended Consequences

**Robert H. Seidman**

**New Hampshire College**

rseidman@minerva.nhc.edu

**ABSTRACT**: "Goals 2000: Educate America Act" aims to, among other things, increase the high school graduation rate to at least 90% and eliminate the graduation rate gap between minority and non-minority students. However well intentioned, this goal is doomed to failure. Powerful systemic forces converge to stabilize the high school graduation rate at about 75% where it has been since 1965 and where no traditional national policy will be able to advance it very much. Even if education policy could succeed in increasing the rate to 90% or beyond, undesirable consequences of potentially great magnitude, especially for the targeted minority groups, would result.

> **Goals 2000: Educate America Act**
>
> **Sec. 102 National Education Goals.**
>
> (2) SCHOOL COMPLETION. --(A) By the year 2000, the high school graduation rate will increase to at least 90 percent. (B) The objectives for this goal are that--
> (i)the Nation must dramatically reduce its school dropout rate, and 75 percent of the students who do drop out will successfully complete a high school degree or its equivalent; and
> (ii) the gap in high school graduation rates between American students from minority backgrounds and their non-minority counterparts will be eliminated.
>
> (Public Law 103-227, 1994)

## I. Introduction

The purpose of the "Goals 2000: Educate America Act" is to promote "coherent, nationwide, systemic education reform." (Public Law 103-227, 20 USC 5801) However well intentioned such an attempt at reform may be, one aspect is doomed to failure. With respect to School Completion (Goal 2), legislators and education policy makers ignore the laws and dynamics of the educational system at their own and our peril.

The "system of education" is a vast and complex enterprise comprising all of the many and different ways society educates it citizens. It is useful to distinguish it from the educational system which possesses a logic and laws of behavior of its own and which can be shown to be highly intractable to attempts to reform it by education policy. This is particularly true with regard to "Goals 2000: Educate America Act."

The theory of the logic and behavior of the educational system illustrates how powerful systemic forces converge to stabilize the high school attainment rate at about 75% where it has been since 1965 and where no traditional national education policy will be able to advance it very much. Even if education policy could succeed in increasing the rate to 90%, or beyond, undesirable consequences of potentially great magnitude, especially for the targeted minority groups, would result.

One undesirable consequence is economic disaster for those who cannot or choose not to complete high school. They will be shut out of important non-educational social benefits (e.g., good job opportunities) unless alternative routes are opened for them. Another consequence is the potential reduction of these very same social benefits for those who do complete high school. A third consequence manifests itself as an unintended, but cruel hoax perpetrated upon the very minorities the Act seeks to help. By virtue of their being the last identifiable group to attain the high school diploma in proportion to their numbers in the age cohort, the high school diploma will not have the same power to secure social goods as it did with previous groups.

Several policy alternatives are explored: 1) push the high school attainment ratio to 100% quickly; 2) reduce the high school attainment rate to the 55-60% level; 3) abandon the normative principle connecting the educational and socioeconomic systems.

- Part II presents a brief outline of a comprehensive and general theory of the logic and behavior of national educational systems (Green, 1980). Certain of its laws and resulting dynamics are exposed.
- Part III presents a non-causal *a priori* aggregate model that illustrates certain systemic dynamics.
- Part IV presents an individual probabilistic utility model that extends the aggregate model. Both models illustrate systemic theory with respect to the Congressional Act and serve to locate critical stages in the growth of the educational system where education policy is most and least effective.
- Part V draws conclusions from the analyses of the two models and discusses several education and non-educational policy alternatives.
- Part VI is an analysis of the results of two models from Raymond Boudon which support the results reported here.
- Appendices A and B contain the mathematics of the Individual Utility model. Appendix C contains the mathematics behind the Aggregate model. Appendix D contains an educational attainment table.(Note 1)

## II. Theory of the Logic and Behavior of the Educational System

A student who leaves school in the middle of the school year in one part of the country and who enters the same grade in a distant part of the country can generally find nearly identical curricula, procedures and facilities. It appears that some sort of system exists.

Education policy is after all, policy for the educational system. But what is the educational system? What are its features? What are the laws of its behavior that set the system in motion? Answers to these questions can help us to assess the potential impact of the

Congressional Act.

*Primary Features.* The primary features of the educational system are threefold:

1. The set of schools and colleges, but not all schools and colleges.
2. These schools and colleges within the system are connected by a medium of exchange which includes those certificates, degrees, diplomas, and the like, that allow one to leave the Nth level of the system in one locality and enter the Nth level in another. They are all instruments by which activities carried out in one place can be recognized and "exchanged" for similar activities of a school or college in some other place. Certain schools and colleges will fall outside of the educational system although they will be within the system of education. Certain proprietary schools may not have their transcripts and diplomas recognized or accepted at other schools that are within the system.
3. The schools and colleges that make up the educational system and that are connected by a medium of exchange are arranged by a principle of sequence: the system of colleges and schools are organized into levels so that if a person has attained (i.e., completed) level N, then he or she has attained level N-1, but not necessarily level N+1.

This principle allows us to speak of persons progressing through the system and seems to be a necessary property of any educational system due in part to differing levels of skill accomplishment, knowledge acquisition and the cognitive development of individuals.

*Secondary Features.* The system also has certain secondary or derivative elements. They are: size, a system of control and a distributive function.

1. Distribution. Every society makes some sort of arrangements for the distribution of its goods (i.e., benefits). The educational system distributes educational goods such as knowledge, skills, and certain kinds of taste, amongst others. In addition to these goods, the system distributes their surrogates, or second-order educational goods such as grades, diplomas, certificates and the like.
2. The derivative element of "control" is less relevant for the present analysis than the others. It turns out that size is of central import since education policy that is effective for one stage of systemic growth may be wholly ineffective at another.
3. System Size. The educational system has eight distinct ways that it can grow (Figure II-1). The present analysis focuses upon "growth in attainment" not only because this is what the Act addresses, but because this mode of growth plays a crucial role in the dynamics of the system which in turn dooms Goal 2 of the Congressional Act to certain failure. (Note 2)

**Figure II-1. The Modes of Growth**

1. The system may expand in response to increases in the school-age population either by increasing the number of units in the system, or by increasing the number of students in the units of the system, or both.
2. Growth in attainment. The system may expand by increasing rates of attendance and survival.
3. Vertical Expansion. The system may expand by adding levels either at the top or at the bottom.
4. Horizontal expansion. The system may expand by assuming responsibility for educational and social functions that are either new, that have been ignored, or that have been carried out by other institutions.
5. Differentiation. The system may expand either by differentiation of programs or institutions or both.
6. Growth in efficiency. The system may expand by intensification, that is, by attempting to do more in the same time or the same in less time.
7. The system may expand by extending the school year or the school day.
8. The system may expand by increasing the number of persons needed to staff it *independently* of the number of students and number of its units, the magnitiude of the school-age population, rates of attendance, survival.

(Green, 1980, p. 10)

There are, however, two more pieces to the system that need to be developed before we can address the notion of growth and size. One is a normative principle connecting the social system with the educational system and the other is the systemic Law of Zero Correlation that relates the strength of the normative principle to system size.

Normative Principle. It is true that some persons, for whatever reason, will come to possess a larger share of educational goods than other persons. This may be due to ability (however it is defined within the system), tenacity, acuity of choice and any number of other reasons.

If non-educational social goods such as income, earnings opportunities and status are distributed by the socioeconomic system on the basis of the distribution by the educational system of educational goods (through the instrumentality of second-order educational goods), then there exists a normative principle that connects the educational and socioeconomic systems.

This normative principle can be rendered as those having a greater share of educational goods merit or deserve a greater share of non-educational social goods. See Figure II-2. The importance and power of this normative principle is, as we shall see, a function of the size of the educational system as measured by the rate of high school attainment. It varies over different stages of systemic growth.

**Figure II-2. Normative Principle Components**

| | | |
|---|---|---|
| Educationally Relevant Attributes | Such as ... | ability, tenacity, acuity of choice, and the like |
| ...are linked to... | | |
| Educational Benefits | Such as ... | knowledge, skills, taste, manners, standards of civility and the like |
| ...are linked to... | | |
| Surrogate Benefits | Such as ... | certificates, diplomas, transcripts, licenses, letters of recommendation, prestige and the like |
| ...are linked to... | | |
| Non-educational Socioeconomic Benefits | Such as ... | income, employment, opportunities, status and the like |

*Law of Zero Correlation.* To understand this law, let us posit a uniform growth curve. Suppose that the educational system grows at a uniform rate over a one hundred year period. That is, there is a uniform increase (10% each decade) in the proportion of each successive age-cohort attaining the 12th level of the system. (The actual growth data is shown in Appendix D.)

When the high school attainment rate is low (e.g., 10% ) the socioeconomic meaning of high school attainment is likely quite negligible. Employers, all things being equal, would have little reason to choose a high school graduate over a non-graduate especially when there are so many of the latter. In the aggregate, high school attainers do not monopolize economic opportunities simply because of attainment. Thus the strength of the normative principle is low. To be a high school drop out when most of your age-cohort drops out presents no serious personal or social problem. See Part A of Figure II-3.

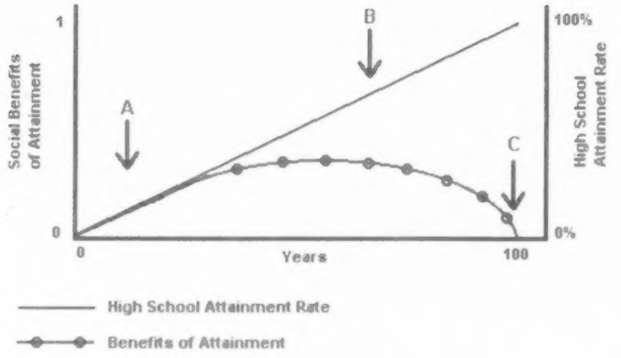

Figure II.3. Uniform Growth Curve and Social Benefits of Attainment

As the size of the educational system increases, the power of the normative principle also increases. Employers now utilize high school attainment as a selection criterion and social goods, such as status and jobs, begin to be preferentially distributed to high school graduates. See Part B of Figure II-3.

However, when the attainment rate reaches 100%, the mere possession of the high school diploma can have no socioeconomic meaning whatsoever. That is, no social goods can be distributed on the basis of high school attainment because everyone has the diploma. It is at this point (and at 0%), that the power of the normative principle is completely destroyed although its power may be weakened well before this point is reached. See Part C of Figure II-3.

The Law of Zero Correlation is a logical tautology. See Figure II-4. It is *a priori* true. For instance, a society could not distribute any of its goods based upon eye color if everyone had the same color eyes. The actual shape of this curve and its inflection points is an empirical matter. However, the models presented here give us some guidance in locating the theoretical inflection points.

| Figure II-4. The Law of Zero Correlation |
|---|
| There is a point of growth of the system at which there is no longer any correlation between educational attainment and either the distribution of educationally relevant attributes in the population or the distribution of non-educational social goods associated with educational attainment.<br>(Green, 1980, p.91) |

*Law of Shifting Benefits and Liabilities.* This is one of the many corollaries of the Law of Zero Correlation. This corollary assures that high school attainment will have a declining social value and that concomitantly, failure to attain the high school diploma will have an increasing social liability, as the attainment rate moves toward the 100% zero correlation point. Thus, as zero correlation is approached, the aggregate social benefits of the attainment group and the aggregate liabilities of non-attainment both increase (Figures II-3 and II-5)

On the liability side, where school leaving was once a possible and viable alternative, it now becomes an evil to be avoided at all costs. These shifting benefits and liabilities make high school attendance and attainment "compulsory" in ways that were surely never meant to be. The personal and social consequences of dropping out of high school can be devastating.

The Law of Shifting Benefits and Liabilities does not specify the points in systemic growth (Sections A, B and C in Figures II-3 and II-5) where the benefits and liabilities of high school attainment shift. However, the two models presented in Parts III and IV do show that when 55% of the 17 year-old age-cohort attains the high school diploma, that group will receive the greater share of social benefits due to the moderate power of the normative principle.

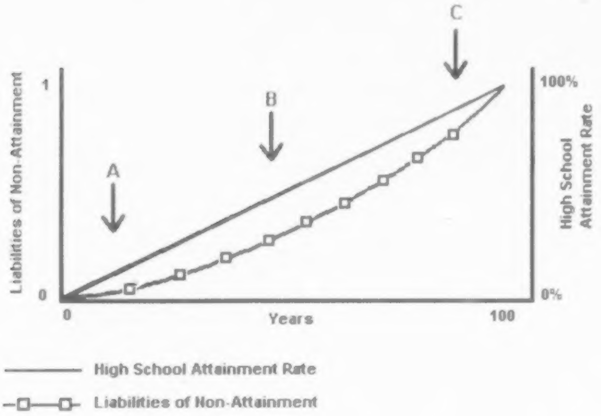

**Figure II-5. Shifting Liabilities of Non-Attainment**

At this point in the growth of the educational system, high school attainment is efficacious in obtaining a disproportionate share of social goods. Thus, a high school diploma becomes a highly sought after good. This corresponds, in the actual growth of the system, to the year 1948. (See Appendix D)

In addition, the models show that when the system becomes fairly large (i.e., 76% high school attainment in 1965), the power of this normative principle begins to decrease even though, historically, the personal and social belief in it remains high. This is prior to zero correlation setting in and may explain why the system has stabilized at around 75% attainment and why it has been so resistant to attemps at education reform.

This is also the point at which the liabilities of non-attainment appear to increase dramatically and where the "drop out problem" became, politically, a problem to be dealt with. Figure II-6 shows the combined effects of the Law of Shifting Benefits and Liabilities and exposes a peculiar paradox: as zero correlation is approached, the aggregate social benefits once associated with high school attainment decline and the associated social liabilities of non-attainment increase.

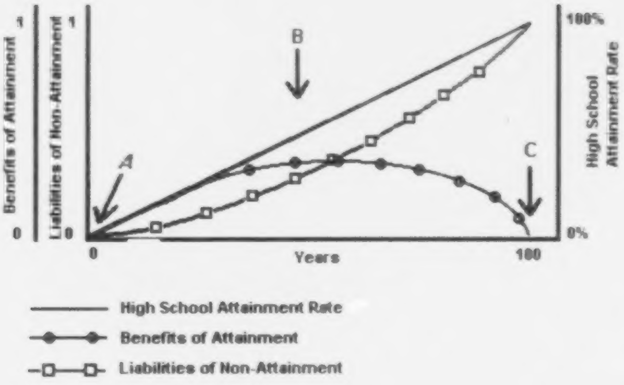

**Figure II-6. Shifting Benefits and Liabilities of Attainment**

If one posits that Section C of Figure II-6 represents the part of the growth of the system where the effects of these laws are maximally felt, then what would befall the minorities that the Congressional Act seeks to help? To address this question, consider two more systemic principles: the Law of Last Entry and the Principle of the Moving Target. These two principles speak to the "Goals 2000" goal of closing the attainment gap (and presumably, the social benefits gap) between minorities and non-minority students.

The Law of Last Entry states that "as we approach the point of universal attainment at any level of the system, the last group to enter and complete that level will be drawn from lower socioeconomic groups." See Figure II-7. However, unlike the Law of Zero Correlation, this law is neither tautological nor *a priori*, but can be considered to be an empirical generalization. The basis for this claim is given in much more detail elsewhere (Green, 1980).

**Figure II-7. The Law of Last Entry**

It appears to be true that no society has been able to expand its total educational enterprise to include the lower status groups *in proportion to their numbers in the population* until the system is "saturated" by the upper and middle status groups. (Green, 1980, p.108)

A corollary of the Law of Last Entry is the Principle of the Moving Target, which states that as the group of last entry reaches its target of proportional 12th grade attainment rate, the target will shift. Note, that if the group of last entry pushes the attainment rate to 100%, then the high school diploma cannot, in and of itself, be used to distribute social benefits to anyone, much less to this last group. Zero correlation will have set in and the target will have shifted to attaining a higher level of the educational system: post-secondary.

However, even if the attainment rate does not reach 100% with the group of last entry (in this case, minority groups), this group will still not reap the same benefits of the high school diploma that previous groups reaped due to the Law of Shifting Benefits and Liabilities. The point in the attainment growth where this occurs is an empirical point. However, the models presented in this paper give us some theoretical guidance.

"Goals 2000" seeks to set and carry out a national policy to increase the high school attainment rate from its present level to at least 90%. If the rate stays below 100%, zero

BEST COPY AVAILABLE

correlation would be avoided. I contend, however, that the effects of merely approaching zero correlation will be felt well before the 90% attainment level is reached (if it ever could be reached!). As the theoretical models which follow show, the felt effect could be one reason why the attainment rate has stabilized for so long at about 75%. Empirical confirmation can be found in (Green, 1980).

## III. THE AGGREGATE MODEL AND APPLICATIONS

*A. The Model* The following Aggregate Model rests upon three idealized assumptions:

1. Non-educational social benefits are always normally distributed in the population under consideration and remain so over time - a change in the high school attainment ratio does not affect the overall normal shape of this distribution;
2. This distribution encompasses those who have attained the high school diploma, but who have not gone on in formal schooling (attainers), and those who have not attained the high school diploma (non-attainers);
3. Society allocates its social benefits in such a way that the attainers monopolize the upper end of the normal distribution.

The first assumption fixes the overall shape of the distribution and offers a particular view of distributed justice. This distribution can be thought to reflect some overall normally distributed attribute or attributes in the total population under consideration. The second and third assumptions tell us that the high school attainers can be found, as a group, lumped at the upper end of the distribution. The third assumption, which admittedly represents an overly rigid meritocratic society, will be altered in the model presented in Part IV.

These three assumptions are realized in Figure III-2, which is a normal distribution in standardized normal form having a grand median (μ) of zero and a standard deviation (ó) of one. Each asymptote is truncated, for computational purposes, at 3.9 standard deviations from the mean. The high school attainment ratio Ø is represented by the shaded area under the curve. This is the proportion of the total population under consideration that has attained the high school diploma. The median value of the social benefits of this group is μ(Ø).

The unshaded portion under the curve is the proportion of the total population that has not attained the high school degree (~Ø) and is equal to (1- Ø). The median value of the social benefit for this group is μ(~Ø).

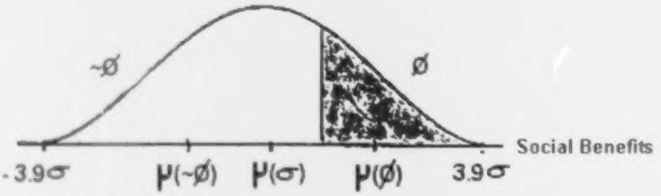

**Figure III-1. Standardized Normal Curve for the Distribution of Social Benefits (Ø = high school attainment ratio; ~Ø = non-attainment ratio; μ = grand median = 0; μ(Ø)= median social benefit for attainer group; μ(~Ø) = median social benefit for non-attainer group; standard deviation = 1)**

### Table III-1

### Median Social Benefits, Their Differences, and Their Rates of Change For Attainer and Non-attainer Groups by High School Attainment Ratio

| (1) Size of Attainment Group: Ø | (2) Attainer Median:μ(Ø) | (3) Non-Attainer Group Median:μ(~Ø) | (4) μ(Ø) - μ(~Ø) | (5) Rate of Change of μ(Ø) | (6) Rate of Change of μ(~Ø) |
|---|---|---|---|---|---|
| 0.01 | 2.575 | -0.012 | 2.587 | . | . |
| 0.05 | 1.960 | -0.063 | 2.023 | 0.2388 | 4.2500 |
| 0.10 | 1.645 | -0.126 | 1.771 | 0.1607 | 1.0000 |
| 0.15 | 1.440 | -0.189 | 1.629 | 0.1246 | 0.5000 |
| 0.20 | 1.283 | -0.253 | 1.536 | 0.1090 | 0.3386 |
| 0.25 | 1.150 | -0.319 | 1.469 | 0.1037 | 0.2609 |
| 0.30 | 1.037 | -0.385 | 1.422 | 0.0983 | 0.2069 |
| 0.35 | 0.935 | -0.454 | 1.389 | 0.0984 | 0.1792 |
| 0.40 | 0.842 | -0.524 | 1.366 | 0.0995 | 0.1542 |
| 0.45 | 0.755 | -0.598 | 1.353 | 0.1033 | 0.1412 |
| 0.50 | 0.675 | -0.675 | 1.350 | 0.1060 | 0.1288 |
| 0.55 | 0.598 | -0.755 | 1.353 | 0.1141 | 0.1185 |
| 0.60 | 0.524 | -0.842 | 1.366 | 0.1237 | 0.1152 |
| 0.65 | 0.454 | -0.935 | 1.389 | 0.1336 | 0.1105 |
| 0.70 | 0.385 | -1.037 | 1.422 | 0.1520 | 0.1091 |
| 0.75 | 0.319 | -1.150 | 1.469 | 0.1714 | 0.1090 |
| 0.80 | 0.253 | -1.283 | 1.536 | 0.2069 | 0.1157 |
| 0.85 | 0.189 | -1.440 | 1.629 | 0.2530 | 0.1224 |
| 0.90 | 0.126 | -1.645 | 1.771 | 0.3333 | 0.1424 |
| 0.95 | 0.063 | -1.960 | 2.023 | 0.5000 | 0.1915 |
| 0.99 | 0.012 | -2.575 | 2.587 | 0.8095 | 0.3138 |

Note that the attainer and non-attainer medians change as a function of the attainment ratio. When the ratio (Ø) is zero, the non-attainer median is equal to the grand median. When the ratio approaches its limit of one, the attainer median approaches the grand median and the non-attainer median approaches -3.9 standard deviations from the grand median. We can easily calculate the values of the attainer and non-attainer medians for different values of the attainment ratio.(Note 3) Table III-1 shows their values, their differences and their rates of change for attainment ratios ranging from 0.01 to 0.99. Figure III-2 is a plot of the attainer and non-attainer medians by the attainment ratio.

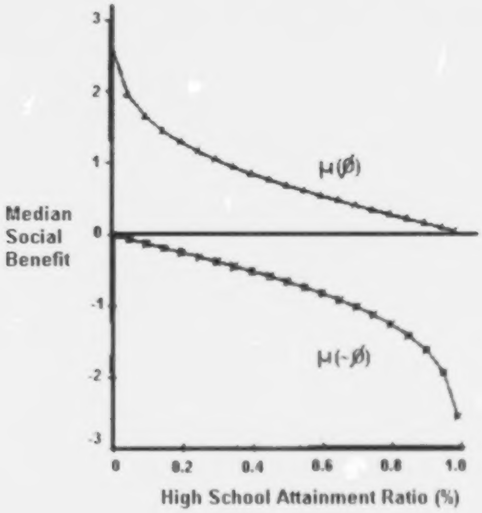

**Figure III-2. Median Social Benefit of Attainer Group (μ(Ø)) and Non-Attainer Group (μ(~Ø)) by High School Attainment Ratio (%) (Ø) (from Table III-1, Columns 2 and 3)**

*B. An Income Disparity Analysis*

A conventional analysis of high school attainer and non-attainer income disparities considers whatever is gained by the attainers to be the magnitude of the liability experienced by the non-attainers. If, for example, the median income of the attainer group is 150% of the non-attainer median income (at a particular attainment ratio), then the benefit to the former group is 50% while the liability to the latter group (in foregone income and earnings opportunities, etc.) is 50%. This approach tends to conceal the full impact of the shifting benefits and liabilities of educational attainment.

Table III-1 and Figure III-1 display another approach to this situation. Here we find the difference between the median benefit of the attainer group and the median benefit of the entire population under consideration (Table III-1, column 2). We do the same for the non-attainer group (Table III-1, column 3). The difference between these two grand-median-dispersions is a measure of the relative position of one group with respect to the other (Table III-1, column 4).

If we think of such social benefits as income, salary and wages, then a conventional supply and demand analysis suggests that as the supply of high school graduates increases, the relative social benefits realized by these graduates, with respect to those with no high school degree, will decline (given a constant market demand for attainers). This is just what happens in the Aggregate Model as the attainment ratio grows from 0.01 to 0.50. However, as the attainment ratio exceeds 50%, the relative advantage of the attainers over the non-attainers increases.(Note 4) See Figure III-2.

These latter results of the model are consistent with certain empirical findings. Time-series U.S. Census data for 18-year-old to 24-year-old males from 1939 (when the national high school attainment ratio was 50%) to 1990 display this phenomenon. (Note 5) A U.S. Senate report which examined the incomes of 24- to 34-year-old males expressed surprise at the "paradox" of increasing relative income for high school attainers over non-attainers. (Note 6)

The interaction between the Law of Zero Correlation and the Law of Shifting Benefits and Liabilities has certain explanatory power when the data are examined as illustrated in the Aggregate Model. The "paradox," cited above, evaporates in light of these systemic dynamics which show the declining benefits associated with attainment and the increasing liabilities associated with non-attainment as the zero correlation point is approached. (Note 7)

*C. Stabilization of the High School Attainment Ratio*

What is the meaning of the "intersection" of the benefit and liability curves in Figure II-6? Although the two curves do not actually intersect (they have different vertical axes), the "intersection" shown in Figure II-6 does illustrate certain interactive systemic effects. This "intersection" can be viewed as an equilibrium point in the growth of the system beyond which it no longer pays (in aggregate social benefit terms) to finish high school but is quite a serious social disaster not to do so. In a way, it is an aggregate recognition of the Law of Zero Correlation and the Law of Shifting Benefits and Liabilities. This phenomenon is illustrated by the Aggregate Model.

Figure III-3 is a plot of the rate of decline of the social benefits of attainment generated by the model. Note that after an attainment ratio of 0.20 the median value declines at a fairly constant rate until the high school attainment ratio reaches 50%. At this point in the growth of the educational system, the rate of decline increases and increases sharply at 75% attainment.

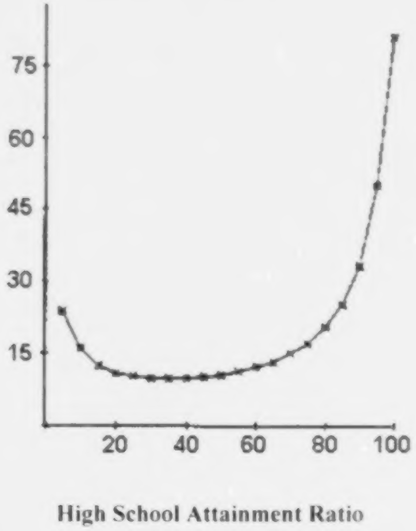

**Figure III-3. Rate of Change of Attainer Group Median (Ordinate) by High School Attainment Ratio (%) (from Table III-1, Column 5)**

Figure III-4 is a plot of the rate of decline of the non-attainer median. Here the median declines at a decreasing rate until 75% attainment at which point the rate begins to increase and then increases sharply at 80% attainment.

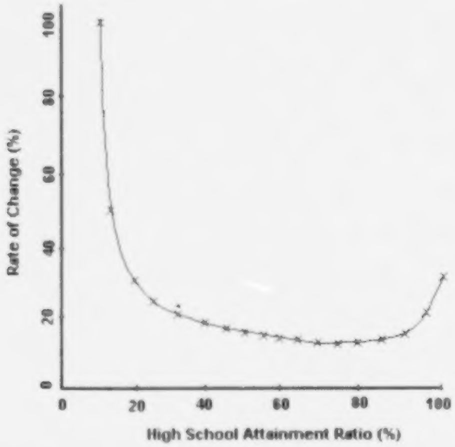

**Figure III-4. Rate of Change of Non-Attainer Group Median by High School Attainment Ratio (from Table III-1, Column 6)**

Thus, the two curves shown in Figure III-2 can be said to contain inflection points which occur in the growth of the system where the high school attainment ratio is about 75%. The stabilization of the national attainment ratio at around 75% may be the social recognition of the phenomenon described by the model.(Note 8)

Is it purely coincidental that the inflection points in the model occur where the national high school attainment ratio has stabilized: at about 75%? Nevertheless, the model does serve to illustrate the phenomenon of systemic "equilibrium" reflecting the interactive dynamics between certain systemic laws. The interaction between these laws offers an account of certain systemic phenomena.

The behavior of the educational system described above is based upon these systemic features: the Principle of Sequence, the distribution of second-order educational goods and the size of the system as measured by the attainment ratio at the twelfth level. Systemic behavior was driven by the power of a logical tautology, its corollary and a normative principle linking the educational and social systems. It is ironic that the "successful" growth of the system, as measured by an increasing high school attainment ratio, appears to sow the seeds of a particularly harsh and peculiar brand of failure. (Note 9)

## IV. THE PROBABILISTIC UTILITY MODEL

The idealized society reflected in the three assumptions underlying the Aggregate Model is a rigidly meritocratic one. By altering the first and third assumptions, (see Section III-A), we can build a model that reflects a society that distributes its non-educational social goods in a somewhat more flexible manner.

Like the Aggregate Model, let us assume that the population under consideration is dichotomized into those who have attained the high school diploma (and nothing beyond it) and those who have not attained the degree. Furthermore, let us assume two independent normal distributions of social goods, one for the attainment group and the other for the non-attainment group. This state of affairs is illustrated in Figure IV-1.

Now let us assume that both of these normal distributions have identical standard deviations. Thus, we can normalize each of the distributions and leave them superimposed, one upon the other, on the social benefits axis. Note that the relative position of the two normal curve means remains unaffected by the standardization (i.e., the standardized and unstandardized means remain stationary). These standardized distributions are shown in Figure IV-1.

### *A. The Standardized Normal Distributions*

Consider the two standardized normal distributions shown in Figure IV-1. Let curves X(Ø) and X(~Ø) represent the distributions of earnings opportunities of high school attainers and non-attainers, respectively. Both curves have their asymptotes truncated, to facilitate the computations to follow, at 3.0 standard deviations above and below their respective means of zero and are superimposed upon a common axis, X, showing an apparent overlap area, E: that area under both curves which has a common X-axis range.

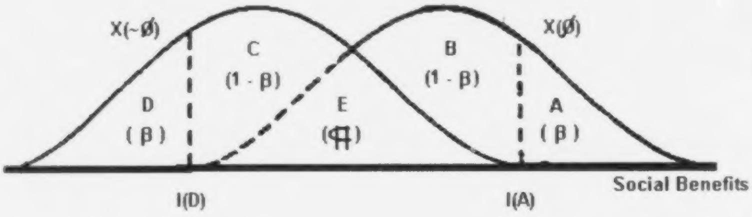

Figure IV-1. Two Overlapping Standardized Normal Curves

We let Ø stand for the ratio of high school attainers to the total population under consideration and let ß stand for the meritocratic parameter. This parameter represents those in the total population, and in particular that proportion of distribution X(Ø) , which monopolizes the highest values of X. It is clear from Figure IV-1 that this parameter imposes an upper-bound on the range of distribution X(~Ø) (i.e., I(A)) and concomitantly places a lower-bound on the range of X(Ø) (i.e., I(D)). Except where ß = 0, the ranges of X(Ø) and X(~Ø) differ.

Let us assume that despite changes in the size of Ø, the original non-standardized normal distributions retain their normal shapes and continue to have identical standard deviations and unchanged means. The X(Ø) mean remains forever fixed and thus for any given Ø , only a change in ß can shift the X(~Ø) curve. A mean/medium analysis of these curves is presented in Appendix B.

Unlike the Aggregate Model, individuals in X(Ø) (i.e., high school attainers) are no longer guaranteed an advantage over persons in X(~Ø) (i.e., non-attainers), with respect to some value of X (level of social benefit). The question now shifts from one of absolute advantage (as in the Aggregate Model) to one of relative advantage. We now ask, what is the probability that an individual will be advantaged with respect to X, over changes in Ø and in ß?

The symbols in Figure IV-1 refer to proportions and are explained in Table IV-1, below.

## Table IV-1
## PROPORTIONAL VALUES OF SECTIONS IN FIGURE IV-1

| Section | Symbol | Meaning |
|---|---|---|
| A | ß | The proportion of the population which is in X(Ø) and which monopolizes the highest X values. This is the value of the meritocratic parameter. |
| B | 1- ß | The proportion of the population which is in X(Ø) and which does not monopolize the highest X values. |
| C | 1- ß | The proportion of the population which is in X(~Ø) and which is not relegated to the lowest X values. |
| D | ß | The proportion of the population which is in X(~Ø) and is relegated to the lowest X values. |
| E | Þ | The area of "intersection" of Section B of X(Ø) and Section C of X(~Ø). |

The above conceptualization allows us to calculate the probabilities of persons falling in any of the five sections of Figure IV-1 as a function of ß and Ø. These probabilities are conditional probabilities of independent events. Table IV-2 gives the formulae for these calculations.

## Table IV-2
## PROBABILITIES

| Section | Probability | Meaning |
|---|---|---|
| A | $\Pr(A \mid X(Ø))=ßØ$ | The probability of residing in Section A is the conditional probability of residing in A given that one already resides in X(Ø). |
| B | $\Pr(B \mid X(Ø))=(1-ß)Ø$ | The probability of residing in Section B is the conditional probability of not residing in A given that one resides in X(Ø). |
| C | $\Pr(C \mid X(\sim Ø))=(1-ß)(1-Ø)$ | The probability of residing in Section C is the conditional probability of not residing in D given that one resides in X(~Ø). |
| D | $\Pr(D \mid X(\sim Ø))=ß(1-Ø)$ | The probability of residing in Section D is the conditional probability of residing in D given that one resides in X(~Ø). |
| E1 | $\Pr(E \mid C \mid X(\sim Ø))=(1-ß)(1-Ø)Þ$ | The probability of residing in Section E given that one is already in X(Ø) , is the conditional probability of residing in E (i.e., Þ) given that one resides in X(~Ø) and resides in Section C. |
| E2 | $\Pr(E \mid B \mid X(Ø))=(1-ß)ØÞ$ | The probability of residing in Section E given that one is already in X(Ø) is the conditional probability of residing in E given that one resides in X(Ø) and resides in B . |

*B. Interpretation of Area E*

The move from proportions in Table IV-1 to probabilities in Table IV-2 is a crucial one. Recall that each distribution represents one part of the dichotomized total population under consideration. The overlapping area E, is not a shared population between the two groups. It simply illustrates the common range of X shared by area B in X(Ø) and C in X(~Ø).

Each person in the total population under consideration has a probability of ending up in one of the two distributions. Since Ø is the proportion of the total population that has attained the twelfth level, any individual has probability Ø of falling under distribution X(Ø) (all other things being equal). Similarly, the probability of not attaining at level 12 is equal to (1- Ø). Of course, Ø + (1- Ø) equals 1.0, which is the total population under consideration. All of this follows from the laws of proportions.

Consider Figure IV-1. As Section A changes in size, X(~Ø) shifts to the left or to the right (recall that we have assumed that changes in Ø do not affect the shape or position of the distributions). The entire area under any one of the two distributions is equal to 1.0. Thus, if ß represents the value of the area of Section A, then 1- ß is the area of Section B. From this we can see that the conditional probability of an individual being an attainer and being a monopolizer of the higher values of X is ßØ.

The laws of symmetry make Section D equal to Section A. Thus, the probability of an individual being a non-attainer and being relegated to the lowest values of X is ß(1- Ø). Similar arguments can be made for Sections B and C. The probabilistic interpretation of Section E is a more complicated matter, however.

Although Sections B and C do not actually have an area in common, they do share the common X-axis range, I(D) to I(A). It is useful to think of Section E as if it is the area of overlap between the two distributions. Recall that the probability of being in C is simply (1-Ø)(1-ß). Now, the probability of being in C and at the same time being within the scope of distribution X(Ø) is just the probability of being in C times the area of Section E. Similarly, the probability of being in B is (1- ß)Ø. The probability of being in B and within the scope of distribution X(~Ø) is just the probability of being in B times the area of Section E.

It should now be clear that Pr(E|C|X(~Ø)) is the probability of any individual non-attainer falling in the same range with and being under the same scope as an attainer. Likewise, Pr(E|B|X(Ø)) is the probability of any individual attainer falling in the same range with and being under the same scope as a non-attainer. These two probabilities need not always be equal. In fact, they are equal only when Ø = 0.50.

What remains is to calculate the area of Section E (i.e.,Þ). This is done in Appendix A.

*C. Results of the Analysis*

Tables IV-3 and IV-4 give the probabilities of falling in Section E given attainment and of falling in Section E given non-attainment, respectively. These Tables are derived from the probability formulae in Table IV-2. To obtain the probabilistic marginal utilities of attainment, we simply perform a matrix subtraction, Table IV-4 minus Table IV-3. The results of this subtraction are shown in Table IV-5.

Note that the marginal utilities decrease for constant Ø and increasing ß, and decrease for constant ß and increasing Ø. Furthermore, each column, reflects about the row where Ø = 0.50 so that each column below this row is the negative converse of the column above.

An inspection of Table IV-5 shows that it is not individually advantageous to obtain the high school diploma until 55% of the population under consideration (17-year old age cohort) does so. The row where Ø =0.50 can be considered to be the indifference level. However, a mean/median analysis shows that, in the aggregate, it is always advantageous to be an attainer rather than a non-attainer. This is so because for all values of ß, μ(Ø) is greater than μ(~Ø) (except when they are equal, when ß = 0). A complete mean/median analysis is given in Appendix B. See columns 4 and 6 in Table B-1.

This analysis of the Probabilistic Utility Model exposes an interesting paradox: in the aggregate it is more advantageous to be an attainer no matter what Ø and ß are; individually this is not always the case. Furthermore, Table IV-5 indicates that the marginal disutility of not attaining the high school degree increases as attainment increases and also increases as the meritocratic parameter decreases! This phenomenon can be vividly seen in the lower left-hand quadrant of Table IV-5.

This quadrant corresponds to the decreasing power of the normative principle as the attainment rate increases toward 100%. As we move from the upper right-hand to the lower left-hand corner on the quandrant diagonal, disutilities can be seen to double, triple and even quadruple at various steps.

Table IV-3
PROBABILITY OF FALLING IN SECTION E GIVEN ATTAINMENT OF LEVEL 12
Meritocratic Parameter (ß)

| * | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.01 | 0.0035 | 0.0022 | 0.0015 | 0.0010 | 0.0007 | 0.0004 | 0.0002 | 0.0001 | 0.0000 | 0.0000 |
| 0.05 | 0.0176 | 0.0112 | 0.0076 | 0.0051 | 0.0033 | 0.0021 | 0.0012 | 0.0005 | 0.0002 | 0.0001 |
| 0.10 | 0.0353 | 0.0224 | 0.0152 | 0.0101 | 0.0067 | 0.0042 | 0.0023 | 0.0011 | 0.0003 | 0.0001 |
| 0.15 | 0.0529 | 0.0336 | 0.0228 | 0.0152 | 0.0100 | 0.0062 | 0.0035 | 0.0016 | 0.0005 | 0.0002 |
| 0.20 | 0.0706 | 0.0448 | 0.0304 | 0.0203 | 0.0134 | 0.0083 | 0.0047 | 0.0022 | 0.0006 | 0.0002 |
| 0.25 | 0.0882 | 0.0560 | 0.0380 | 0.0254 | 0.0167 | 0.0104 | 0.0058 | 0.0027 | 0.0008 | 0.0003 |
| 0.30 | 0.1058 | 0.0672 | 0.0456 | 0.0304 | 0.0200 | 0.0125 | 0.0070 | 0.0033 | 0.0010 | 0.0003 |
| 0.35 | 0.1235 | 0.0785 | 0.0532 | 0.0355 | 0.0234 | 0.0146 | 0.0082 | 0.0038 | 0.0011 | 0.0004 |
| 0.40 | 0.1411 | 0.0897 | 0.0608 | 0.0406 | 0.0267 | 0.0166 | 0.0094 | 0.0044 | 0.0013 | 0.0004 |
| 0.45 | 0.1588 | 0.1009 | 0.0684 | 0.0456 | 0.0301 | 0.0187 | 0.0105 | 0.0049 | 0.0014 | 0.0005 |
| 0.50 | 0.1764 | 0.1121 | 0.0759 | 0.0507 | 0.0334 | 0.0208 | 0.0117 | 0.0055 | 0.0016 | 0.0005 |
| 0.55 | 0.1940 | 0.1233 | 0.0835 | 0.0558 | 0.0367 | 0.0229 | 0.0129 | 0.0060 | 0.0018 | 0.0006 |
| 0.60 | 0.2117 | 0.1345 | 0.0911 | 0.0608 | 0.0401 | 0.0250 | 0.0140 | 0.0066 | 0.0019 | 0.0006 |
| 0.65 | 0.2293 | 0.1457 | 0.0987 | 0.0659 | 0.0434 | 0.0270 | 0.0152 | 0.0071 | 0.0021 | 0.0007 |
| 0.70 | 0.2470 | 0.1569 | 0.1063 | 0.0716 | 0.0468 | 0.0291 | 0.0164 | 0.0077 | 0.0023 | 0.0007 |
| 0.75 | 0.2646 | 0.1681 | 0.1139 | 0.0760 | 0.0501 | 0.0312 | 0.0176 | 0.0082 | 0.0024 | 0.0008 |
| 0.80 | 0.2822 | 0.1793 | 0.1215 | 0.0811 | 0.0534 | 0.0333 | 0.0187 | 0.0088 | 0.0026 | 0.0008 |
| 0.85 | 0.2999 | 0.1905 | 0.1291 | 0.0862 | 0.0568 | 0.0354 | 0.0199 | 0.0093 | 0.0027 | 0.0009 |
| 0.90 | 0.3175 | 0.2017 | 0.1367 | 0.0913 | 0.0601 | 0.0374 | 0.0211 | 0.0099 | 0.0029 | 0.0009 |
| 0.95 | 0.3352 | 0.2130 | 0.1443 | 0.0963 | 0.0635 | 0.0395 | 0.0222 | 0.0104 | 0.0031 | 0.0010 |

* Proportion of 12th Level Attainers

Table IV-4
PROBABILITY OF FALLING IN SECTION E GIVEN ATTAINMENT BELOW LEVEL 12
Meritocratic Parameter (ß)

| * | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.01 | 0.3493 | 0.2219 | 0.1504 | 0.1004 | 0.0661 | 0.0412 | 0.0232 | 0.0109 | 0.0032 | 0.0010 |
| 0.05 | 0.3352 | 0.2130 | 0.1443 | 0.0963 | 0.0635 | 0.0395 | 0.0222 | 0.0104 | 0.0031 | 0.0010 |
| 0.10 | 0.3175 | 0.2017 | 0.1367 | 0.0913 | 0.0601 | 0.0374 | 0.0211 | 0.0099 | 0.0029 | 0.0009 |
| 0.15 | 0.2999 | 0.1905 | 0.1291 | 0.0862 | 0.0568 | 0.0354 | 0.0199 | 0.0093 | 0.0027 | 0.0009 |
| 0.20 | 0.2822 | 0.1793 | 0.1215 | 0.0811 | 0.0534 | 0.0333 | 0.0187 | 0.0088 | 0.0026 | 0.0008 |
| 0.25 | 0.2646 | 0.1681 | 0.1139 | 0.0760 | 0.0501 | 0.0312 | 0.0176 | 0.0082 | 0.0024 | 0.0008 |
| 0.30 | 0.2470 | 0.1569 | 0.1063 | 0.0710 | 0.0468 | 0.0291 | 0.0164 | 0.0077 | 0.0023 | 0.0007 |
| 0.35 | 0.2293 | 0.1457 | 0.0987 | 0.0659 | 0.0434 | 0.0270 | 0.0152 | 0.0071 | 0.0021 | 0.0007 |
| 0.40 | 0.2117 | 0.1345 | 0.0911 | 0.0608 | 0.0401 | 0.0250 | 0.0140 | 0.0066 | 0.0019 | 0.0006 |
| 0.45 | 0.1940 | 0.1233 | 0.0835 | 0.0558 | 0.0367 | 0.0229 | 0.0129 | 0.0060 | 0.0018 | 0.0006 |
| 0.50 | 0.1764 | 0.1121 | 0.0759 | 0.0507 | 0.0334 | 0.0208 | 0.0117 | 0.0055 | 0.0016 | 0.0005 |
| 0.55 | 0.1588 | 0.1009 | 0.0684 | 0.0456 | 0.0301 | 0.0187 | 0.0105 | 0.0049 | 0.0014 | 0.0005 |
| 0.60 | 0.1411 | 0.0897 | 0.0608 | 0.0406 | 0.0267 | 0.0166 | 0.0094 | 0.0044 | 0.0013 | 0.0004 |
| 0.65 | 0.1235 | 0.0785 | 0.0532 | 0.0355 | 0.0234 | 0.0146 | 0.0082 | 0.0038 | 0.0011 | 0.0004 |
| 0.70 | 0.1058 | 0.0672 | 0.0456 | 0.0304 | 0.0200 | 0.0125 | 0.0070 | 0.0033 | 0.0010 | 0.0003 |
| 0.75 | 0.0882 | 0.0560 | 0.0380 | 0.0254 | 0.0167 | 0.0104 | 0.0058 | 0.0027 | 0.0008 | 0.0003 |
| 0.80 | 0.0706 | 0.0448 | 0.0304 | 0.0203 | 0.0134 | 0.0003 | 0.0047 | 0.0022 | 0.0006 | 0.0002 |
| 0.85 | 0.0529 | 0.0336 | 0.0228 | 0.0152 | 0.0100 | 0.0062 | 0.0035 | 0.0016 | 0.0005 | 0.0002 |
| 0.90 | 0.0353 | 0.0224 | 0.0152 | 0.0101 | 0.0067 | 0.0042 | 0.0023 | 0.0011 | 0.0003 | 0.0001 |
| 0.95 | 0.0176 | 0.0112 | 0.0076 | 0.0051 | 0.0033 | 0.0021 | 0.0012 | 0.0050 | 0.0002 | 0.0001 |

* Proportion of 12th Level Attainers

Table IV-5
PROBABILISTIC MARGINAL UTILITIES OF ATTAINMENT OF LEVEL 12
Meritocratic Parameter (ß)

| * | 0.10 | 0.20 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.01 | 0.3460 | 0.2200 | 0.1490 | 0.0990 | 0.0650 | 0.0410 | 0.0230 | 0.0110 | 0.0030 | 0.0010 |
| 0.05 | 0.3180 | 0.2020 | 0.1370 | 0.0910 | 0.0600 | 0.0370 | 0.0210 | 0.0100 | 0.0030 | 0.0010 |
| 0.10 | 0.2820 | 0.1790 | 0.1220 | 0.0810 | 0.0530 | 0.0330 | 0.0190 | 0.0090 | 0.0030 | 0.0010 |
| 0.15 | 0.2470 | 0.1570 | 0.1060 | 0.0710 | 0.0470 | 0.0290 | 0.0160 | 0.0080 | 0.0020 | 0.0010 |
| 0.20 | 0.2120 | 0.1340 | 0.0910 | 0.0610 | 0.0400 | 0.0250 | 0.0140 | 0.0070 | 0.0020 | 0.0010 |
| 0.25 | 0.1760 | 0.1120 | 0.0760 | 0.0510 | 0.0330 | 0.0210 | 0.0120 | 0.0050 | 0.0020 | 0.0010 |
| 0.30 | 0.1410 | 0.0900 | 0.0610 | 0.0410 | 0.0270 | 0.0170 | 0.0090 | 0.0040 | 0.0010 | 0.0000 |
| 0.35 | 0.1060 | 0.0670 | 0.0460 | 0.0300 | 0.0200 | 0.0120 | 0.0070 | 0.0030 | 0.0010 | 0.0000 |
| 0.40 | 0.0710 | 0.0450 | 0.0300 | 0.0200 | 0.0130 | 0.0080 | 0.0050 | 0.0020 | 0.0010 | 0.0000 |
| 0.45 | 0.0350 | 0.0220 | 0.0150 | 0.0100 | 0.0070 | 0.0040 | 0.0020 | 0.0010 | 0.0000 | 0.0000 |
| 0.50 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 0.55 | -0.0350 | -0.0220 | -0.0150 | -0.0100 | -0.0070 | -0.0040 | -0.0020 | -0.0010 | 0.0000 | 0.0000 |
| 0.60 | -0.0710 | -0.0450 | -0.0300 | -0.0200 | -0.0130 | -0.0080 | -0.0050 | -0.0020 | -0.0010 | 0.0000 |
| 0.65 | -0.1060 | -0.0670 | -0.0460 | -0.0300 | -0.0200 | -0.0120 | -0.0070 | -0.0030 | -0.0010 | 0.0000 |
| 0.70 | -0.1410 | -0.0900 | -0.0610 | -0.0410 | -0.0270 | -0.0170 | -0.0090 | -0.0040 | -0.0010 | 0.0000 |
| 0.75 | -0.1760 | -0.1120 | -0.0760 | -0.0510 | -0.0330 | -0.0210 | -0.0120 | -0.0050 | -0.0020 | -0.0010 |
| 0.80 | -0.2120 | -0.1340 | -0.0910 | -0.0610 | -0.0400 | -0.0250 | -0.0140 | -0.0070 | -0.0020 | -0.0010 |
| 0.85 | -0.2470 | -0.1570 | -0.1060 | -0.0710 | -0.0470 | -0.0290 | -0.0160 | -0.0080 | -0.0020 | -0.0010 |
| 0.90 | -0.2820 | -0.1790 | -0.1220 | -0.0810 | -0.0530 | -0.0330 | -0.0190 | -0.0090 | -0.0030 | -0.0010 |
| 0.95 | -0.3180 | -0.2020 | -0.1370 | -0.0910 | -0.0600 | -0.0370 | -0.0210 | -0.0100 | -0.0030 | -0.0010 |

* Proportion of 12th Level Attainers

## V. RESULTS, CONCLUSIONS, CONJECTURES and POLICY ALTERNATIVES

These models illustrate the theoretical limitations of education policy designed to increase the high school attainment rate to 90% or above and to help minorities share in the "benefits" of educational attainment. They are formal models and are not grounded in empirical results. Like Raymond Boudon's models (see Part VI), they avoid the cross-sectional and variable confounding of survey data. They illustrate the power of a logical tautology in conjunction with a normative principle. However, these idealized models are not without limitations

The Aggregate Model seems, on the face of it, too meritocratic for our present society. The distribution of social benefits may not in reality, be normal and their means (as shown in the Utility Model) may not remain constant with systemic growth (which is clearly not the case in the Aggregate Model). Nonetheless, these models can serve as "benchmarks" against which to compare other logico-mathematical models containing different assumptions, and still others based upon empirically derived data. They also add to our database of models.

*Policy Alternatives.* The results of the models developed in this analysis suggest a number of possible alternative education policy scenarios. Three such follow.

- Push the High School Attainment Rate to 100% quickly.
  Given that attempts to reduce social inequalities by increasing the national high school attainment ratio will fail, what would be the consequences of entirely eliminating educational attainment inequality at the high school level? That is, push the high school attainment rate to 100% so that the high school diploma can no longer be the basis for the distribution of non-educational social goods.
  This approach has two major pitfalls. First, the system had better reach 100% attainment

very quickly so as to minimize the hardships that will have to be endured by the ever decreasing percentage of non-attainers. Second, even if such a result could be achieved, the original inequality problems would remain unsolved since the problems would merely be shifted to the next higher level of the educational system - postsecondary.
If the normative principle persists (and there is no reason to assume that it will not) then the distributional instrument of social goods will shift to the postsecondary level. This level is, for the most part, selective. One does not only choose to go on, one is chosen. Thus, enormous pressures will come to bear upon this level to alter its selectivity feature. One can argue that this pressure is already fairly strong.

- Reduce the High School Attainment Rate to the 55-60% Level.
  This level is below the "equilibrium point" of the Aggregate Model and close to the "indifference" level of the utility model. This is the point at which the effects of the decline in the social benefits of attainment and the precipitous rise in the social liabilities of non-attainment are (theoretically) thought to begin.
  Of course, careful consideration needs to be given to the provision of ample opportunities for all to continue their education (i.e., pursue learning). Such a policy must avoid an inequitable distribution of the non-attainers based on educationally irrelevant attributes such as race, class and ethnic background. Admittedly, a policy of this sort would not enjoy widespread political support.
- Abandon the Normative Principle. The two previous alternatives assumed the continued presence of the normative principle. But what would life be like without it? The abandonment of this principle might be the most efficacious, but a politically and socially difficult, way to reduce educational and socioeconomic inequality.
  If educational attainment is no longer used as an instrument for the distribution of non-educational social goods, then perhaps education could once again be pursued for the benefits that are intrinsic in the educational goods themselves and not for the socioeconomic advantages that disappear and reappear with ever increasing rates and different levels of attainment.
  Such a move might signal the end of the illusion that the educational system is a solution to practically every social ill. I do not claim to know just what new instruments for the distribution of social benefits would arise, nor how one could go about judging their desirability as a replacement for educational attainment. However, a reconsideration of the socioeconomic normative principle that disproportionately rewards formal educational attainment might prove to be a beneficial exercise.

## VI. ANALYTICAL POSTSCRIPT: BOUDON'S MODELS OF INEQUALITY OF EDUCATIONAL and SOCIAL OPPORTUNITY

Two models created by the French Sociologist, Raymond Boudon (Boudon, 1974) support the results of the two models presented here. Boudon's models are of inequality of educational opportunity (IEO) and inequality of social opportunity (ISO). He analyses their relationship to one another and to the educational and social systems. Some of Boudon's relevant results and analyses follow.

### *A. Boudon's IEO and ISO Models and the Theory of Educational Systems*

Boudon's models and his analyses are highly suggestive in many ways. In addition to a methodological approach which avoids some of the pitfalls of factorial analysis (i.e., partial accounting for total variance, cross-sectional "illusions," and lack of quantitatively adequate data), Boudon adds an important dimension to the description of the normative behavior of the type of educational system spawned in Western industrial societies. This dimension, system animation, is of fundamental import in helping to provide a clear and precise picture of the dynamics of systemic motion.

By observing (and modeling) the over-time cumulative effects of the various factors affecting the educational system's growth, Boudon is able to discern the logical limits and consequences of this growth. The ceiling-effect and the exponential mechanism that combine to drive the IEO model help generate a number of observations and paradoxes that bear significantly upon the theory of educational systems as presented here.

### Some Familiar Paradoxes

One of the most striking paradoxes generated by Boudon's models is that "other things being equal" (which is seldom the case), educational growth has the effect of increasing social and economic inequality. This happens even when the system becomes more egalitarian with respect to educational opportunity (EO).

This paradox rests upon the assumption that income is dependent upon educational attainment level. Over time, educational level and socioeconomic status increase with educational level increasing more rapidly the higher the socioeconomic level. Since both of these factors are "independently" responsible for income differentials, "economic inequality will increase over time along with social inequality, for the latter is correlated with the former." (Boudon, 1974, pg. 188)

The paradox is completed when we add another important conclusion reached by the application of Boudon's model: change in social stratification is the only factor that can substantially affect the model's exponential mechanism and hence ISO. This leads Boudon to conclude that educational growth can partially explain the "persistence of economic inequality in Western societies." (ibid., 188) It is quite remarkable that Boudon's model and the models pesented here reach identical conclusions using such different but complementary methods.

### The Success-Breeds-Futility Paradox

Another paradox illustrates just how the apparent success of the educational system leads to futility for some participants and how the system fuels the fires of its own expansion. Boudon's models indicate that one of the main endogenous factors responsible for the increase in educational demand is the over-time change in the status expectations of individuals with respect to educational level.

> ...as time goes on, the structure of expectations associated with the two highest levels of education is constant; intermediate levels are affected most adversely; the structure of expectations relating to the lowest levels of education becomes less favorable, too, but it is less influenced by the overall educational increase than are the intermediary levels. (ibid., 149)

Thus, as IEO decreases over time and the educational system expands at all levels, the social status expectations for persons at intermediate educational levels decrease and these persons must raise their levels just to maintain constant social status expectations. This treadmill effect means that while the relation between educational level and social status changes very little over time, the number of years of schooling associated with each of the educational levels increases.

Thus, while the average level of educational attainment in the population increases, the educational levels that are associated with particular status expectations are "simultaneously moving upward." As individuals demand more and more education over time, the individual return tends to be nil, while the aggregate return on this demand is high. The lower socioeconomic classes are compelled to demand more education (especially if the higher classes to do), for not to do so condemns these lower classes to constantly falling social status expectations. However, more educational demand only retards this diminution in status and does not increase the lower class's chances of achieving increased social status.

This is a particularly frustrating paradox, for in a meritocratic society where the normative principle holds, an individual seems to have an advantage in securing as much education as he or she can. However, when many individuals seek additional education, the aggregate effects of this demand decrease the social status expectations associated with most of the educational levels. This causes people to demand even more education in the next time period.

This paradox lends support to a number of results due to the interactions between various systemic principles such as the Law of Zero Correlation, the Principle of Shifting Benefits and Liabilities, the Law of Last Entry, and the Principle of the Moving Target. Boudon shows that when expectations associated with some particular educational level become reduced, a decrease in expectations at all levels results. (ibid., Table 8.4, 147)

Boudon sees evidence that this point has been reached at the secondary level in some industrial societies, but "it seems that not even the most advanced industrial societies have achieved a proportion of college students so large that a severe decrease in the expectations at this level can be observed." (ibid., 150) One wonders whether or not the American educational system has moved to a point beyond Boudon's claim? Because of their logico-mathematical nature, the models presented here are generalizable over all systemic levels. Already, over 60% of the high school graduates enter higher educational institutions (National Center for Education Statistics, 1994). It may not be long before the system approachs zero correlation at this level!

Perhaps in anticipation of zero correlation at the college level, Thurow has called for a "system of post-secondary education for the non-college bound student" (Thurow, 1994). However, I suggest that such a "system" (even if established independently of the educational system) would itself be absorbed into the educational system and therefore be subject to its laws and thus perpetuate the paradoxes discussed here. Such is the power of the dynamics of the educational system.(Note 10)

### *B. Further Observations on Systemic Growth*

While the paradoxes generated by Boudon's model are important for establishing the boundaries and limitations of educational systems, there are other observations on growth that warrant exploration.

Boudon, in his Appendix to Chapter 9, indicates that by manipulating the demand for education (i.e., predicating demand in the educational system upon exogenous rather than endogenous factors), equality of educational opportunity (EEO) can be affected. This is the only alternative, other than changing social stratification, that he offers to remedy IEO and ISO.

Now, if the number of positions (student slots) in the educational system at the highest level remains unchanged and if the number of positions at the middle level is increased by D during time period t to t+1, and if the number of positions at the lowest level is decreased by D during this same time period -- then, how is the number of persons with lowest social background T(t ) who reach at least the middle educational level affected by the value of D?

Boudon concludes on the basis of this "modified" model that T(t) is an increasing function of time and an increasing function of D. Furthermore, T(t) increases at a decreasing rate as a function of process-phase. According to Boudon, the duration's of the three phases are a function of D ("an increase in D has the effect of shortening the first and second phases..."). Thus non-linear returns in T(t) are associated with increase in the value of D. This thesis is presented in expanded form in (Boudon, 1976).

This "modified" model (reflecting an "ideal-typical planned educational system") results in a decrease in IEO through the manipulation of demand, while the IEO parameter, "a", remains constant over time. (This IEO parameter has marked similarities to the meritocratic parameter, ß, presented in the Aggregate and Individual Utility models.) The free-market endogenous educational system creates what appear to be insurmountable problems (i.e., the paradoxes).

On the other hand, the exogenous educational system, permits us in theory at least, to correct some of these undesirable effects. Boudon rightfully questions the high social costs of this remedy. Nevertheless, this "modified" model may provide additional insights into the growth mechanism of the system and may have enormous implications for policy and planning especially if the demand for education is to be controlled. It deserves further study.

### *C. A Logistic Growth Curve*

In an intriguing footnote (ibid., 201, ff.3), Boudon suggests that in conjunction with the paradoxes cited above, there is a particular point in the free-market educational system development where "growth is more rapid at the higher level than at the secondary level and thus a decrease in IEO and ISO is curtained." (ibid., 199) This growth, fueled by unrestrained demand for more education, may lead to a state of "latent crisis." This runaway exponential growth trend may be checked by a "braking process" that is proportional to the trend, leading to a logistic rather than an exponential growth curve.

What are the circumstances that would lead to this braking process and would these circumstances be endogenous or exogenous to the educational system? The answers to these questions are fundamental to education policy. These answers appear to be intimately related to many of the systemic principles in the general theory of educational systems.

Finally, what is to be made of Boudon's enigmatic statement that "the concern of all industrial societies with short-term higher education can be better understood in the light of the dialectic between the exponential growth of educational demand and the (proportional) braking process...?" Perhaps the theory of the educational system and the models put forward here can shed some light on this question?

## Notes

1. Originally presented at the Annual Meeting of the American Educational Research Association San Francisco, California April 19, 1995 Session 16.36 The Political Context of Educational Reform (Division G; SIG/Politics of Education). Some of the ideas and models presented here have appeared in various forms and stages of completion in previous works. In particular: Thomas F. Green(1980) with David P. Ericson and Robert H. Seidman; Seidman(1982); Seidman(1981). The analysis of Boudon's work has never been published.

2. This paper uses the high school attainment rate as the measure of systemic "size due to growth in attainment." One reason is that this is what the Congressional Act focuses on. Another, is that the 12th grade is the last level of the educational system that is non-selective. For the most part, one not only chooses to go on to post-secondary education, one is chosen. It is this fact, together with certain systemic laws, that illustrates the inherent futility of certain education policies at particular stages of systemic growth.
I use the 17 year-old age-cohort to measure the high school attainment rate. This is the cohort used by the National Center for Education Statistics (1995) to track the high school attainment ratio since 1869. The models presented here are based upon a dichotomized population: those who have not completed high school and those who have but have not gone on to the post-secondary level of the system.
However, some researchers use a different age-cohort. For example, the National Education Goals Panel uses the 19-20 year-old age cohort (National Education Goals Panel, September 1994). Other studies report high school completion rates amongst various age cohorts, including 21-22 year-olds and even 29-30 year-olds (National Center for Education Statistics, 1993).The numeric ratios will differ, of course. A standard measure of high school "completion and school leaving" has been proposed. The "appropriate unit of analysis" is the graduating class cohort.(Hartzell, 1992).

3. A sample calculation can be found in Appendix C.

4. It is probably unreasonable to apply the model at the lower attainment rates where the power of the normative principle is very low. However, the model does serve to illustrate the idea that the relative benefit disparity between the two groups first decreases and then increases. This phenomenon suggests that a particular educational policy appropriate for one stage of systemic growth may not be appropriate for another.

5. U.S. Bureau of the Census, Decenial Census Reports for 1940, 1950, 1960, 1970; Current Population Reports, P60, nos. 85, 90, 92, 97, 101. U. S. Bureau of the Census, Current Population Reports for 1984, 1987, 1990; P70, nos. 11, 21, 32 ("Educational Background and Economic Status").

6. See Levin(1972) for a traditional analysis of the relevant data.

7. For an extended analysis from another methodological perspective, see Appendix C in (Green, 1980).

8. See the Table reproduced in Appendix D (National Center for Education Statistics, 1995). It is interesting to note that the U.S. Government projection of the high school attainment ratio to the year 2006 keeps it at about 74% (using the 18 year-old cohort). Why? No reason is given. See Tables 26 and B4 (National Center for Education Statistics, 1996).

9. This irony (in the form of paradoxes) is addressed by Boudon (1974) and is analyzed in Part VI above. Boudon's models confirm the results of the Aggregate and Individual models.

10. For an example of such an absorption scenario, see Seidman's (1982) analysis of the "lifelong learning system."

## REFERENCES

Boudon, Raymond (1974) Education, Opportunity, and Social Responsibility. NY: Wiley.

Boudon, Raymond (1976) "Elements pour une theorie formelle de la mobilite sociale" Quality and Quantity, 5; pp. 39-85.

Green, T.F., with Ericson, D.P. and Seidman, R.H., (1980). Predicting the Behavior of the Educational System. NY: Syracuse University Press.

Hartzell, G., McKay, J. & Frymier, J. (1992), "Calculating Dropout Rates Locally and Nationally with the Holding Power Index." ERIC ED 343 953.

Levin, Henry M. et al. (1972) The Cost to the Nation of Inadequate Education. A Report Prepared for the Select Committee on Equal Educational Opportunity of the U.S. Senate. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics (1993) Dropout Rates in the United States: 1992. Washington, DC: U.S. Government Printing Office. (ERIC ED 363 671)

National Center for Education Statistics (1996) Projections of Education Statistics to 2004. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics (1995) Digest of Education Statistics. Washington, DC: U.S. Government Printing Office.

National Education Goals Panel (September 1994) "National Education Goals Report: Building a Nation of Learners" Washington, DC: U.S. Government Printing Office.

Public Law 103-227 (1994), "Goals 2000: Educate America Act". 103d Congress.

Seidman, R.H. (1981) "The Explanatory Power of Two Idealized Models of Educational and Social Attainment." American Research Association Annual Meeting, Los Angeles, CA. (ERIC: ED 206068)

Seidman, R.H. (1982). The Logic and Behavioral Principles of Educational Systems: Social Independence or Dependence?, chapter in M.S. Archer (Ed.), The Sociology of Educational Expansion: Take-off, Growth and Inflation in Educational Systems (SAGE Studies in International Sociology 27), CA: Sage.

Thurow, Lester (1994) "Education and Falling Wages," New England Journal of Public Policy, 10(1), pp. 53-5.

U.S. Bureau of the Census, Decenial Census Reports (1940, 1950, 1960, 1970) Current Population Reports; P60 nos. 85, 90, 92, 97, 101. Washington DC: U.S. Government Printing

Office.

U.S. Bureau of the Census (1940, 1950, 1960, 1970) Current Population Reports: P60 nos. 85, 90, 92, 97, 101. Washington D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census (1984, 1987, 1990) Current Population Reports: Educational Background and Economic Status; P70 nos. 11, 21, 32. Washington DC: U.S. Government Printing Office.

---

# APPENDIX A

## CALCULATIONS OF SECTION E AREA

To calculate Þ, we begin by truncating the asymptotes of the two standardized normal curves (Figure IV-1) at 3.0 standard deviations above and below their respective means. As a result, we lose 0.26% of the population of any one curve.

Since the two curves are identical (i.e., both are standardized normal curves), the point on the X-axis (μ(I) directly below the point of intersection, I) lies midway between the X(Ø) and X(~Ø) distribution means, μ(Ø) and μ(~Ø), respectively. This follows from the laws of symmetry, since Section D is always equal to Section A in area. Figure A-1 emphasizes the area of intersection in Figure IV-1.

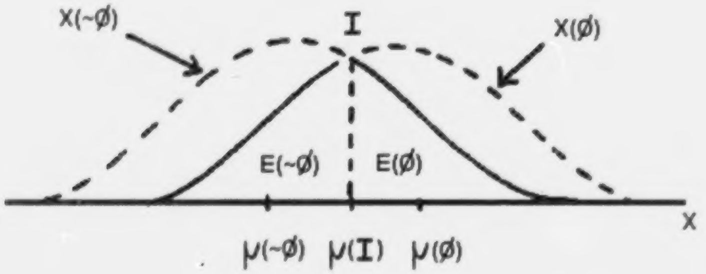

**Figure A-1. Section E Area Emphasized**
**(E(Ø) and E(~Ø) correspond to E1 and E2 , respectively, in Table IV-2)**

We know by symmetry, that the area to the right of the vertical line Iu(I) to μ(I) on curve X(~Ø) (i.e., area E(~Ø)) is equal to the area to the left of line I to μ(I) on curve X(Ø) (i.e., area E(Ø) ). Thus, twice E(~Ø) or twice E(Ø) gives us Þ, the area of Section E(Ø) .

Now we can proceed to develop a pair of algorithms that enable us to calculate area E(~Ø).

The area Þ, equals 1.0 when ß equals zero. In this situation, X(~Ø) and X(Ø) are superimposed one upon the other. Since μ(~Ø) = μ(Ø) , their relative difference, ¶, is equal to the absolute value of μ(~Ø) - μ(Ø) which is equal to zero. When ß =1.0, area Þ equals zero. In this case, X(~Ø) and X(Ø) are mutually exclusive and ¶ equals 6.0. Between these two extremes, ß ranges from zero to 1.0.

We first examine the case where ß ranges from zero to 0.5 and then the case where it ranges from 0.5 to 1.0. (Note that 0.5 is used throughout as an approximation to 0.4987, which is used in the calculations due to truncation.)

### CASE 1: (0 < = ß = > 0.5)

Consider Figure A-2. The relative distance, ¶, between the two means, µ(~Ø) and µ(Ø) , is equal to the distance on the X-axis under area A (i.e., the area corresponding to the value of ß).

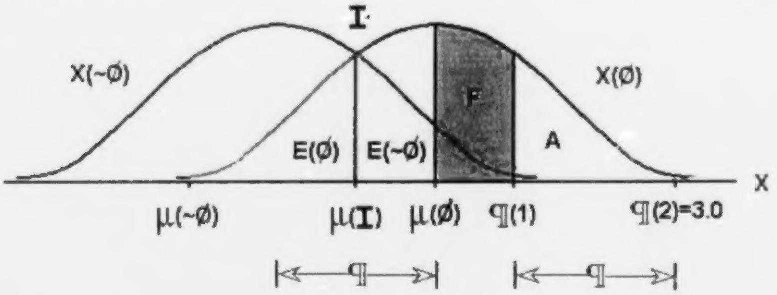

**Figure A-2. Case 1: Where ß Ranges from 0 to 0.5**

Note that when ß = 0, the two means, µ(~Ø) and µ(Ø), coincide simply because the two curves, X(~Ø) and X(Ø) , are superimposed one upon the other. As the value of ß increases, the X(~Ø) curve is shifted to the left, a distance equal to the distance on the X-axis under Section A. Call this distance ¶, which is the value of the X(~Ø) curve translation.

Since ¶(2) = 3.0, we need only find ¶(1) in order to find ¶ (i.e., ¶ = ¶(2) - ¶(1) ). Area F is equal to 0.4987 - G and ¶(1) is found from a standardized normal curve table. Once we have computed ¶, we can locate µ(I) with respect to µ(~Ø) . See Figure A-3.

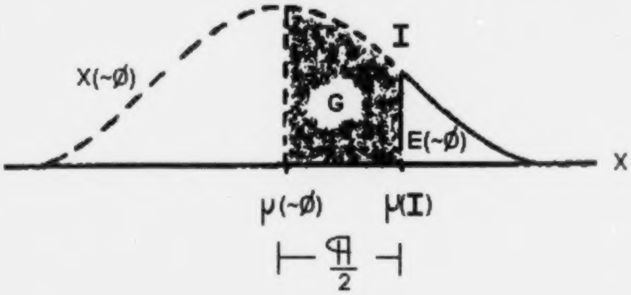

**Figure A-3. The Parameters for Finding ß**

Note that µ(I) lies ¶/2 above µ(~Ø) . Area G is found from a standardized normal curve table. Area E(~Ø) is equal to 0.4987 - G. The area ¶, is simply twice area E(~Ø) . The algorithm for this computation is shown in Algorithm A-1.

## ALGORITHM A-1

### CASE 1: WHERE ß RANGES FROM 0 TO 0.5
(Refer to Figures A-2 and A-3)

Step

1. F = 0.4987 - ß
2. ¶(1) from standardized normal curve table
3. ¶ = ¶(2) - ¶(1)
4. µ(I) = ¶/2 with respect to µ(~Ø)
5. G from standardized normal curve table
6. E(~Ø) = 0.4987 - G

7. Þ = 2(E(~Ø)) )

**CASE 2: (0.5 < = ß = > 1.0)**

Figure A-4 depicts the situation for this case, and the algorithm for the computation of Þ is shown in Algorithm A-2.

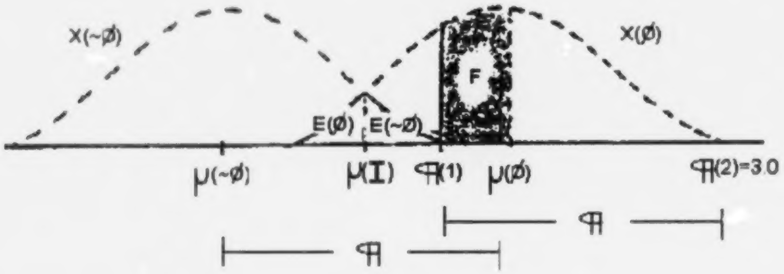

**Figure A-4. Case 2: Where ß Ranges from 0.5 to 1.0**

## ALGORITHM A-2

### CASE 2: WHERE ß RANGES FROM 0.5 TO 1.0
### (Refer to Figures A-3 & A-4)

Step

1. F = ß - 0.4987
2. ¶(1) from standardized normal curve table
3. ¶ = ¶(2) + ¶(1)
4. μ(I) = ¶/2 with respect to μ(~Ø)
5. G from standardized normal curve table
6. E(~Ø) = 0. 4987 - G
7. Þ = 2(E(~Ø)) )

Table A-1, gives the values of Þ for ß values in steps of 0.1. Table A-2 gives the intermediate values of F, ¶(1) , ¶, μ(I) , G, μ(~Ø) for ß values in steps of 0.1.

## Table A-1

### VALUES OF Þ AS A FUNCTION OF ß

| ß | Þ |
|---|---|
| 0 | 1.0000 |
| 0.10 | 0.3872 |
| 0.20 | 0.2776 |
| 0.30 | 0.2124 |
| 0.40 | 0.1666 |
| 0.50 | 0.1310 |
| 0.60 | 0.1006 |
| 0.70 | 0.0750 |
| 0.80 | 0.0516 |
| 0.90 | 0.0294 |
| 0.95 | 0.0172 |
| 1.00 | 0 |

## Table A-2

### INTERMEDIATE VALUES FROM ALGORITHMS A-1 and A-2

| ß | F | ¶(1) | ¶ | μ(I) | G | E(~Ø) |
|---|---|---|---|---|---|---|
| 0 | . | . | 0 | . | . | . |
| 0.10 | 0.3987 | 1.275 | 1.725 | 0.8625 | 0.3051 | 0.1936 |
| 0.20 | 0.2987 | 0.835 | 2.165 | 1.0825 | 0.3599 | 0.1388 |
| 0.30 | 0.1987 | 0.520 | 2.480 | 1.2400 | 0.3925 | 0.1062 |
| 0.40 | 0.0987 | 0.250 | 2.750 | 1.3750 | 0.4154 | 0.0833 |
| 0.50 | 0 | 0 | 3.000 | 1.5000 | 0.4332 | 0.0655 |
| 0.60 | 0.1013 | 0.255 | 3.255 | 1.6275 | 0.4484 | 0.0503 |
| 0.70 | 0.2013 | 0.530 | 3.530 | 1.7650 | 0.4612 | 0.0375 |
| 0.80 | 0.3013 | 0.850 | 3.850 | 1.9250 | 0.4729 | 0.0258 |
| 0.90 | 0.4013 | 1.290 | 4.290 | 2.1450 | 0.4840 | 0.0147 |
| 0.95 | 0.4513 | 1.660 | 4.660 | 2.3300 | 0.4901 | 0.0086 |
| 1 | . | . | 6.000 | . | . | . |

## APPENDIX B

### MEAN/MEDIAN ANALYSIS OF THE PROBABILISTIC UTILITY MODEL

We can set the Model in motion. See Figure B-1. Note that when ß = 0, the following equalities hold:

1. μ(B) = μ(C) = μ(I) = μ(~Ø) = μ(Ø)
2. Absolute value of (μ(A) - μ(I)) = absolute value of (μ(D) - μ(I))

   When ß = 1, another set of equalities hold:

3. μ(C) = μ(B) = μ(I)
4. μ(A) = μ(~Ø)
5. μ(D) = μ(~Ø)
6. Absolute value of (μ(A) - μ(I)) = absolute value of (μ(D) - μ(I))

Between these two extremes, it is possible to calculate the relative differences between medians (μ(Ø) and μ(~Ø) are the grand means and grand medians of their respective distributions) of the various sections of the two curves shown in Figure B-1.

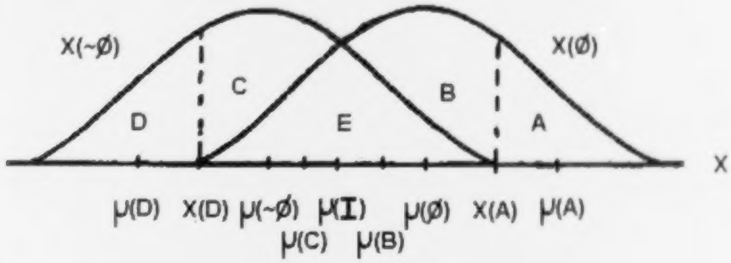

**Figure B-1. Medians/Means for Sections of Curves**

Assume that μ(Ø) remains constant and that both curves retain their normal shapes as the size of Ø (and concomitantly, ~Ø) and ß change. We take μ(Ø) as our point of reference, since it remains constant, and calculate the other medians with respect to it.

1. Schema's for Median Calculations for Changing Values of ß

We begin, as we did in Appendix A, by truncating the asymptotes of the two standardized normal curves at 3.0 standard deviations above and below their respective means. Medians μ(A) and μ(B) have already been calculated in the Aggregate Model and can be found in columns 2 and 3 of Table III-1.

μ(~Ø) is the distance on the X-axis under Section A. This distance is the ¶ value computed as an intermediate step by Algorithms 1 and 2. See Table A-2. μ(I) is simply one half μ(~Ø) and is also computed as an intermediate step by Algorithms 1 and 2. See Table A-2.

We now develop schemas that compute the values of μ(C) and μ(D) , for changing values of ß.

Due to the symmetry of the two curves and the equality of Sections A and D, median μ(C) will always be as much to the right of μ(~Ø) as μ(B) is to the left of μ(Ø) . Thus,

(7) μ(C) = μ(~Ø) - μ(B). (7)

In a similar fashion, μ(D) will always be as much to the left of μ(~Ø) as μ(A) is to the right of μ(Ø) . Thus,

(8) μ(D) = μ(~Ø) - μ(A). (8)

Table B-1 displays the results of these computations.

2. Changing Means (μ(Ø) and μ(~Ø) ) With Changing Ø and Constant ß.

We have assumed throughout that the size of Ø has no effect on the means of the dichotomized populations. Furthermore, for computational purposes, we have assumed that only μ(~Ø) was affected by changing ß and that μ(Ø) remains permanently anchored.

It is not unreasonable to assume that both means change with changing Ø and that both means change with changing ß. However, both of these cases reduce to the analysis that has already been performed for the probability distributions generated by the formulae in Table IV-2 (constant µ(Ø) for changing Ø and changing ß).

## Table B-1

### INTERMEDIATE VALUES FROM ALGORITHMS 1 AND 2

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|
| ß | µ(Ø) | µ(A) | µ(B) | µ(I) | µ(C) | µ(~Ø) | µ(D) |
| 0 | 0 | 3.0 | 0 | 0 | 0 | 0 | -3.0 |
| 0.10 | 0 | 1.645 | -0.126 | -0.8625 | -1.5990 | -1.725 | -3.370 |
| 0.20 | 0 | 1.283 | -0.253 | -1.0825 | -1.9120 | -2.165 | -3.448 |
| 0.30 | 0 | 1.037 | -0.385 | -1.2400 | -2.0950 | -2.480 | -3.517 |
| 0.40 | 0 | 0.842 | -0.524 | -1.3750 | -2.2260 | -2.750 | -3.592 |
| 0.50 | 0 | 0.675 | -0.675 | -1.5000 | -2.3250 | -3.000 | -3.675 |
| 0.60 | 0 | 0.524 | -0.842 | -1.6275 | -2.4130 | -3.255 | -3.779 |
| 0.70 | 0 | 0.385 | -1.037 | -1.7650 | -2.4930 | -3.530 | -3.915 |
| 0.80 | 0 | 0.253 | -1.283 | -1.9250 | -2.5670 | -3.850 | -4.103 |
| 0.90 | 0 | 0.126 | -1.645 | -2.1450 | -2.6450 | -4.290 | -4.416 |
| 0.95 | 0 | 0.063 | -1.960 | -2.3300 | -2.7000 | -4.660 | -4.723 |
| 1.0 | 0 | 0 | -3.0 | -3.0 | -3.0 | -6.0 | -6.0 |

To construct the probability tables for changing means, we can use the probability distributions generated by the formulae in Table IV-2. We need only know the sizes of Ø and ß, and the relative difference between the two dichotomized population means (see Appendix A). This relative difference, absolute value of µ(Ø) - µ(~Ø) , is a function only of the size of ß. Thus, if both means change with changing Ø and with changing ß, and if we know the relative difference between the means, we can calculate the new ß. We can then consult the existing probability tables produced by the formulae in Table IV-2.

3. Non-normal Distributions with Equal and Unequal Ranges

The same sort of mean/median and probability analyses that have been performed for normal distributions can be performed for non-normal distributions. One must, however, first derive the formulae for the various curves and utilize the calculus to obtain the areas in questions and their shifting means and medians. The mathematics involved in this kind of analysis is more complex.

## APPENDIX C

### A SAMPLE CALCULATION FOR THE AGGREGATE MODEL

Here is a sample calculation of the median value of the social benefits for high school attainers and non-attainers.

Suppose that the attainment ratio stands at 30 percent. See Figure C-1. We know that the attainer group monopolizes the social benefits ranging in value from 0.52 to 3.9 standard deviations from the grand mean.

The median benefit for this group is thus µ(Ø) = 1.037 standard deviations. This is the point under the Ø portion of the total distribution where half of the high school attainers (i.e., 15 percent) lie to the right and where the other half lie to the left.

The median social benefits for the remaining 70 percent of the total population (i.e., the non-attainer group) is $\mu(\sim Ø) = -0.385$ . This is the point under the Ø portion of the total distribution where one half of the high school non-attainers (i.e., 35 percent) lie to the right and the other half lie to the left.

The median social benefit values are derived from the standardized normal distribution, which represents a particular normal distribution of social benefits. If it turns out that, for this particular normal distribution, the median of the total distribution is $8,000 with a standard deviation of $2,500, we can easily calculate the medians (in dollars) of the attainer and non-attainer groups.

Attainer Group Median: $10,593 = $8,000 + (1.037 x $2,500); non-Attainer Group Median: $7,038 = $8,000 + (-0.385 x $2,500).

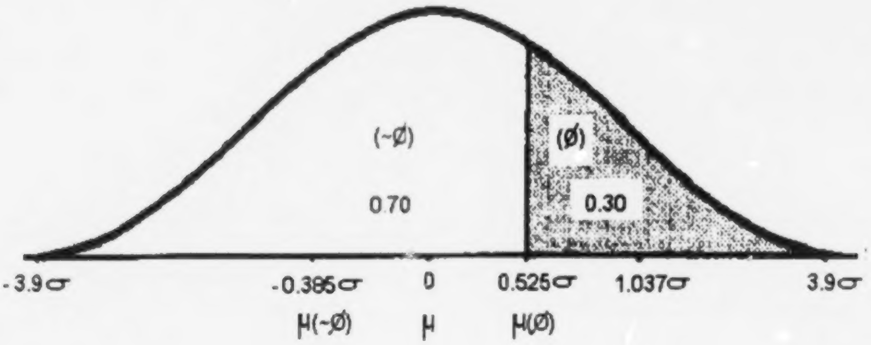

Figure C-1. Standardized Normal Curve for the Distribution of Social Benefits

( Ø= high school attainment ratio; ~Ø= non-attainment ratio; grand median=0; $\mu(Ø)$ = median social benefit for attainer group; $\mu(\sim Ø)$= median social benefit for non-attainer group; standard deviation = 1)

It is probably unreasonable to apply the model at the lower attainment ratios where the power of the normative principle is very low. However, the model does serve to illustrate the idea that the relative benefit disparity between the two groups first decreases and then increases. This phenomenon suggests that a particular education policy appropriate for one stage of systemic growth might not be appropriate for another stage.

## Appendix D

### Empirical High School Attainment Data*

| School Year | Graduates as Percent of 17-year-old Population |
|---|---|
| 1869-70 | 2.0 |
| 1879-80 | 2.5 |
| 1889-90 | 3.5 |
| 1899-00 | 6.4 |

| | |
|---|---|
| 1909-10 | 8.8 |
| 1919-20 | 16.8 |
| 1929-30 | 29.0 |
| 1939-40 | 50.8 |
| 1947-48 | 52.6 |
| 1949-50 | 59.0 |
| 1951-52 | 57.4 |
| 1953-54 | 59.8 |
| 1955-56 | 63.1 |
| 1956-57 | 63.1 |
| 1957-58 | 64.8 |
| 1958-59 | 66.2 |
| 1959-60 | 69.5 |
| 1960-61 | 67.9 |
| 1961-62 | 69.3 |
| 1962-63 | 70.9 |
| 1963-64 | 76.7 |
| 1964-65 | 72.1 |
| 1965-66 | 76.4 |
| 1966-67 | 76.3 |
| 1967-68 | 76.3 |
| 1968-69 | 77.1 |
| 1969-70 | 76.9 |
| 1970-71 | 75.9 |
| 1971-72 | 75.6 |
| 1972-73 | 75.0 |
| 1973-74 | 74.4 |
| 1974-75 | 73.6 |
| 1975-76 | 73.7 |
| 1976-77 | 73.8 |
| 1977-78 | 73.0 |

| | |
|---|---|
| 1978-79 | 71.7 |
| 1979-80 | 71.4 |
| 1980-81 | 71.7 |
| 1981-82 | 72.4 |
| 1982-83 | 72.9 |
| 1983-84 | 73.1 |
| 1984-85 | 72.4 |
| 1985-86 | 72.0 |
| 1986-87 | 71.8 |
| 1987-88 | 72.1 |
| 1988-89 | 71.0 |
| 1989-90 | 72.4 |
| 1990-91 | 73.2 |
| 1991-92 | 73.1 |
| 1992-93 | 73.2 |
| 1993-94 | 73.1 |

* National Center for Education Statistics (1995). Table 98 in *Digest of Education Statistics.* Washington, DC: U.S. Government Printing Office.

## About the Author

**Robert H. Seidman**

**New Hampshire College**

*rseidman@minerva.nhc.edu*
2500 N. River Rd., Manchester, NH 03106-1045 USA
Voice: 603-644-3102 X3346
FAX: 603-644-3150
Robert H. Seidman is a professor at the New Hampshire College Graduate School and the Executive Editor of the Journal of Educational Computing Research.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@zodiac.rutgers.edu*

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Sherman Dorn
*dornsj@ctrvax.vanderbilt.edu*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 12** **August 8, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS
ARCHIVES.Permission is hereby granted to copy any
article provided that EDUCATION POLICY ANALYSIS
ARCHIVES is credited and copies are not sold.

---

# Public School Reform: Potential Lessons from the Truly Departed

**J. Dan Marshall**
**Pennsylvania State University**

*jdm13@psuvm.psu.edu*

**James P. Valle**
**Donegal School District (PA)**

**Abstract:**
In this article, the authors present data from a small study of 19 families who educate their children at home in rural Pennsylvania. Findings relative to why they opted out of the public education system and whether they would return are analyzed in light of a previously established construct (Idealogue/Pedagogue) before being used to critique and expand it in light of broader cultural concerns. The authors argue, overall, that home educators are asserting their historical option of cultural agency and schooling. (Note 1)

If "school reform" is a bandwagon, then the parade is still in progress. Most of the grand proposals earlier composed by politicians, pundits, policy wonks, and professors have evolved into smaller, more locally pertinent endeavors by actual change participants (educators, students, parents and community members). In the worst case, the continuing accumulation of school reform efforts is understood as succeeding waves of perpetual hassle and silliness which disturb the basic soundness of business-as-usual. In the best case, such efforts become a representation of participants' commitment to the repetitive nature of the learning process: desiring to know and understand - acting upon these desires - making sense of and reflecting upon those actions - identifying new or different desires to know and understand. Thus, in the best case, school reform efforts should be here to stay.

Those who care about examining and acting upon the quality of their local schools seek information from numerous sources, including their own experiences, outside consultants, beliefs and opinions collected from local, state, and national polls, and "the literature" of academia. But they seldom tap the one segment of their community which may provide the most unique perspective: parents who have opted out of the local public school system. We suspect that this group -- particularly those families who have taken it upon themselves to provide education at home -- may have something important to offer those working to change public education. In this article, we discuss our preliminary foray into the lives of several

Pennsylvania home educators in light of public school reform efforts.

## Home Education -- A Return to Educational Agency

The philosopher Jane Roland Martin (1996) recently discussed the relationship between a nation's cultural wealth and its commitments to education in the broadest sense. Working from the premise that cultural wealth must be broadly defined to include multiple "conceptions of high, popular, and material culture, and . . . countless other items as well" (p. 6), she suggests that the educational responsibility or agency for transmitting this wealth must return to the breadth it once enjoyed. And for a good deal of time in our history the home bore much of this educational agency.

Prior to the great American experiment of educating all young people in publicly funded schools, most families bore primary responsibility for the education of their children. Support for these efforts in the form of reinforcement, refinement, and reorientation could be counted on from the community, extended family, and the church. While schools existed in our colonial period, they had little to offer the majority of people and little currency as a stand alone educational site. Even during the nineteenth century, the "common school" movement was accompanied by corresponding community located educational efforts (public libraries, agricultural societies, etc.). Slowly, beginning with Massachusetts in 1852 and ending with Mississippi in 1918, the United States became a land of compulsory schooling laws which, Supreme Court decisions in the early 1920s notwithstanding, legitimized schools as the primary educational agency. "It was only in the 20th century," Martin writes, "that schools came to be seen as the sum total of education" (1996, p. 8).

Martin's (1996) overarching point is that "the assets that our culture has placed in school's keep [i.e., preparing young people for their places in the world of politics, work, and the professions] represent one small portion of the [cultural] wealth" of our country (p.8); much of our remaining cultural wealth (largely that which pertains to popular and material culture) was assigned to the educational agency of home. Over time, the primacy of schools as bearers of educational agency and transmitters of dominant, high cultural wealth has overwhelmed the educational agency of the home and its historically gendered role in preserving other forms of cultural wealth.

Social and political activities blossoming in the 1960s helped to tie these "other forms of cultural wealth" directly to public schooling. As the federal government moved into the business of national curriculum development, activists and parents raised questions about the overall relevance of schooling to students' "real lives." The growing movements around people's rights (collective and individual) combined with a deteriorating political environment to produce a general desire to among many to question authority. Humanistic and critical thinking and practices complicated public schools which were caught in the throes of desegregation, while values -- ranging from religious and spiritual to democratic and political -- were noted as absent from the overall school experience. At the same time, new alternatives to the business-as-usual of public schooling began to appear.

The late John Holt embodies the transitional spirit of school reform during these times. From his call for sweeping changes in public schools in 1964 (How Children Fail) he came to believe that parents and families, themselves, must re-take control of their children's education. With the establishment of his magazine, Growing Without School in 1977, Holt dedicated the rest of his life to nurturing and supporting the civic-minded educational agency of the home by popularizing home education (Marshall and Sears, 1985).

"Home schooling," the more popular term to describe families who teach their children at home (Litcher & Schmidt, 1991)(Note 2), has grown from roughly 15,000 to 350,000 students within the past ten years (Jeub, 1994; Lines, 1991). While in 1980 only three states had established laws to permit and control home schooling, 34 states have done so to date. Pennsylvania's more liberally enabling home schooling legislation (unanimously passed by both legislative bodies) went into effect in late 1988, following the state's supreme court ruling on the unconstitutionality of its previously confining statute (Klicka, 1990).

We have a long-term interest in learning more about the pedagogical practices and guiding beliefs of these Pennsylvania home educators. In the following section we describe our initial effort to establish lines of communication and develop a sense of their feelings toward education at home and in schools. Perspectives from Pennsylvania Home Educators

Following the passage of this more liberal Pennsylvania legislation, one of us (Jim) became involved with home educators as the "District Evaluator" of their efforts. In addition to his work as an elementary school teacher, his evaluator's job is to see that home-based educational activities concur with the law's requirements. Jim seems a wise choice for this role in that he is a former administrator of a Christian school, a longstanding member of the community, and (alongside his wife) a home educator himself. No less important, perhaps, is his reputation throughout the community as a vocal supporter of home education. When requested, Jim also serves families in the role of "independent evaluator" (an advocate who is personally selected by each home education family) to certify that the family's efforts have been "appropriate" in the eyes of the law. These roles provide him with "official" (though not necessarily intimidating) access to home educators in several school districts, including his own.

Jim's local school district includes about 15,000 people and can be rightfully described as largely rural and conservative. The county's picturesque landscape in southeastern Pennsylvania, once dominated by neatly spaced barns and silos, is increasingly dappled with housing developments -- up from 49 new housing permits in 1980 to 518 in 1990. Most of the district's 2,508 students begin school in one of four elementary buildings, move on to the lone middle school, and eventually matriculate to the central high school.

During the present school year some 55 children from this district are being educated at home -- a number that has risen steadily since 1988. We wondered what has prompted so many families to sidestep the public school system and take on the work of educating their students at home. How might they characterize their motivation for and commitment to the educational agency they have regained as home educators?

As the first step in a larger study designed to explore the curricular understandings and practices of home educators, we contacted all 27 home education families from Jim's district, along with 16 additional families for whom he serves as independent evaluator (a total of 43 families). Each family received a personal letter from Jim, describing and seeking their participation in the larger study, and asking them to complete and return a brief (one side of one page) survey designed to collect preliminary demographic information (number of school-aged children, number of years residing in district, etc.) along with answers to two simple questions. Those considering further participation signed these forms and provided telephone numbers; others remained anonymous.

Nineteen families (44%) responded to our initial inquiry -- a response rate we accepted as adequate for our exploratory purposes, given that many home educators prefer not to interact with interlopers (Clark, 1994). They raise an average of three school-aged children, all of whom are home educated in 15 of these families. Respondents have been Pennsylvanians for an average of more than 23 years (range of 1-45) and have lived within their particular school district for an average of 10 years. On average, these families have been conducting home education for nearly five years, though they range in this work from one to 11 years.

*Compelling Reasons for Home Education.* Our survey made two simple, straightforward requests: 1) to describe the most compelling reason(s) for home education and 2) to say whether or not public schooling might again become an option and, if so, under what conditions. In cases where families offered more than one response, we identified their first one as a "primary" response, followed by a "secondary" response, etc.

Our home education families offered at least five different reasons which compel them to teach their children at home. Though recorded by respondents as such, these reasons may not be mutually exclusive. Here, we present them separately.

The least often mentioned reason was "cost." Only three of the 19 families identified home education as a choice resulting from the prohibitive cost of private schooling, though none of these saw cost as a primary reason. These three families identify themselves as having chosen home education for religious reasons as well.

Five respondents specified what we call "family cohesion" as a compelling (though not primary in any case) reason for home education. Here, respondents speak of benefits like "family unity," and "spending time together." These families have been conducting home education from four to nine years, and all who listed family cohesion also identified themselves as religiously motivated home educators.

Some 36% of families (seven) named "peer influence" as a compelling reason for leaving (or never entering) the public schools. This reason, typically expressed as "influences of other students" such as "boy-girl relationships," "drugs, sex, alcohol," and "becoming part of the Tin crowd," cut across the range of respondents in most respects (number of years doing home education, primary reasons for home education, etc.). While only two of those identifying "peer influence" as a compelling reason for home education also included religious reasons, "peer influence" was the sole, primary, or secondary reason noted by all who included it.

Fewer than half (8) of our respondents explicitly stated religious beliefs as a compelling reason for home education, with six of these eight families listing this as their sole or most compelling reason. Representative of such beliefs would be the following statement: "We home school so that our children might receive an education that is consistent with our belief that God created the world and is in control of it." Interestingly, all but two of these families have been home educating for five or more years (the upper end of our range).

Within our sample, the most frequently offered reason for educating children at home pertains to the problematic quality of life and learning found in public schools -- what we call "learning concerns." These concerns ranged from dull academic environments to an over-emphasis on college-bound students; from inappropriate labeling of children to an inability to individualize instruction; from teachers who don't care to administrators "out to get" certain problem kids. Thirteen of our 19 families (68%) found such matters compelling, with seven listing learning concerns as either their sole or primary reason for abandoning public schools. Though this reason was identified by families who have been practicing home-based education from 1-7 years, it is the dominant (i.e., sole or primary) reason among those seven responding families with the fewest (1-3) years of practice in home education.

Among these 19 families, 58% (eleven) identified multiple reasons compelling them to separate themselves from the district's public schools. Six of these eleven families include their religious beliefs as one of those reasons (almost all as a primary or secondary reason), yet only three of those six families list both religious convictions and learning-related concerns (in contrast, for example, to "family cohesion" which is mentioned by five of these six families). Of those eight families who offered but a single compelling reason for electing home education, two were religious and one was peer influence; the remaining five noted "learning concerns."

*Returning to the Public School Fold.* When asked whether or not they would "ever consider" returning to public schools and if yes, why, the answer from nearly 75% of our respondents was simply "No." Within this group of parents, seven were unequivocal and emphatic; three would do so only as a result of some personal catastrophe (e.g., illness or death); two would consider such a move only if their children requested it; one would return children to public schools only if the law required it; and one family would consider public schooling again only if the schools somehow changed.

The remaining five families were clearly less strident in their feelings about a possible return to public schools. Two families are among only four from our sample who simultaneously have children attending public schools and, we suspect, see public schools as a viable place for some of their children but not others. In the remaining three cases, one family may consider returning their child to the public schools in order to take advantage of a senior high school vocational-technical career training option, another is considering a return in light of their local school's apparently more enlightened understanding of their child's particular needs (in this case, "hyperactivity"), and the third would consider a return if they felt they were unable to adequately prepare their children for post-high school learning.

Looking at the question differently, nearly 60% of these home educators take the position that nothing short of personal catastrophe or the long arm of the law would get their children back into public schools. Of this group, eight have been practicing home education for five years or more. None of those who have abandoned public schools for religious reasons would return to the public schools, nor would six of the nine families who included learning concerns but not religious beliefs among their reasons to educate their children at home.

The five families that would consider returning their children to the public school fold all say that they left (or decided against ever enrolling in the first place) due to concerns about their children's learning and/or peer influence. All but one of these families have been home educating for three years or less, and all respond to this question with respect to their children.

That is, for these families, home education seems to be a choice which has been made in the best interests of (and perhaps in consultation with) their school-aged children. This group of parents, it seems, will "see how it goes" -- for their children at home and with respect to what's happening within their neighborhood public schools.

## Ideologues, Pedagogues and Beyond

In light of the extant scholarship on "home schooling," none of this is especially new. Numerous studies have surfaced similar motivating factors (see, for example, Mayberry, 1989; Mayberry & Knowles, 1989), though most find much more significance in the religion factor than we presently do (Lines, 1991). Much of this work has been built on a scaffold developed by Jane Van Galen (1988, 1991) who characterizes parents who teach their children at home as falling into "two broad categories" of home education parents: Ideologues and Pedagogues. Acknowledging "tremendous variation" within and across these groupings, Van Galen (1988) describes Ideologues as those parents, largely conservative Christian in their religious beliefs, who "object to what they believe is being taught in public and private schools and . . . seek to strengthen their relationship with their children." In contrast, Pedagogues believe that "schools teach whatever they teach ineptly" and that, based on their respect for their children's intelligence and creativity, "children learn best when pedagogy taps into the child's innate desire to learn." Thus, Ideologues abandon public schools when they feel that schools teach "a curriculum that directly contradict[s] their own values and beliefs," while Pedagogues opt for home education "because they [believe] that their children would be harmed academically and emotionally by the organization and pedagogy of formal schools" (Van Galen, 1988, p. 55).

In some respects, Van Galen's categories seem to fit our preliminary inquiry. Those Pennsylvanians we contacted who home educate for "religious" reasons are the same parents who identified "family cohesion" and "prohibitive cost" (each of the three families mentioned Christian schools here) as compelling reasons for sustaining their home education efforts. Thus, we could refer to this collection of eight families as similar to Van Galen's Ideologues. These families constitute the more veteran home schoolers among our respondents -- with half of them pre-dating Pennsylvania's 1988 home education law. Further, while only two families within this group listed religious beliefs as their sole compelling reason for home education, six of the 11 families offering multiple reasons could be characterized as Ideologues. All of this suggests that while religious beliefs may be strong among this group, the concomitant benefit of family cohesion along with the prohibitive cost of private Christian schools help to keep them educating children at home. Only three of these eight families, for example, specifically offered any sort of "learning concern" as a compelling reason for leaving or never even considering the public schools.

Van Galen's "Pedagogue" category also finds strong support from our preliminary findings. With the exception of the three families who listed both religious beliefs (Ideologues) and learning concerns (Pedagogues) as compelling reasons for dismissing public schools, our Pedagogues do, indeed, seem to highlight concerns about academic and/or emotional harm resulting from "the organization and pedagogy of formal schools." Further, this group was unmistakably more willing than their Ideologue counterparts to consider returning their children to public schools under certain circumstances.

What we find problematic about this categorization scheme, however, is its temptation to allow us to reduce what Harris & Fields (1982) call this "outlaw generation" of parents into easily identifiable (and thus, easily disposable) caricatures: Ideologues become right-wing Christian fanatics and Pedagogues become New Age eco-progressives. In short, we risk distancing "them" from "us."

Marginalizing home educators as "them" further serves to support and sustain all the myths which have grown up around this movement -- including myths about who "can" teach, what does and doesn't get taught/learned, and the social isolation of home-educated students (Meighan, 1984). Again, much available information indicates otherwise (see, for example, Calvery and Others, 1992; Frost, 1988; Groover & Endsley, 1988; Ray, 1988; Ray & Wartes, 1991; Stough, 1992; Tipton, 1990; Webb, 1989).

More importantly, however, such myths reinforce the primacy of school as the sole educational agency, particularly when they are perpetuated by professional educators like education professor Robert Slywester, who believes that "Home-schooled children miss important opportunities," and Thomas Shannon, executive director of the National School Boards Association, who believes that "Few [home educating] parents . . . are objectively qualified to do so" (Cohen, 1995, p.7; see, also, Mahan & Ware, 1987).

But exploring and explaining these myths detours our attention to larger and more important matters concerning educational agency and civic-minded public schooling. Arguing that only schools can provide social competence or state certified teachers sidesteps the larger and more immediate questions pertaining to which specific civic and cultural responsibilities belong to and might best be accomplished within schools and how those differ from responsibilities which belong to and might best be addressed within the home and family.

Home and school -- the two primary sites of educational agency -- must, Jane Roland Martin argues, begin to balance and share responsibilities for maintaining our cultural wealth. As Martin puts it:

> It is downright irrational to persist in assigning school a function that is defined in relation to and relies on home's educational agency while denying the existence of that very agency. It is also the height of folly to assign what we take to be our one and only educational agent the task of preparing children for life in the public sphere . . . Besides, given the great changes home has undergoing in recent decades and the importance to both the development of children and the life of society of the cultural wealth that home has been charged with transmitting, to equate education with schooling, yet continue to endorse a function for school that is premised on home's carrying out an opposite but equally important function, is short-sighted in the extreme. (1996, p.9)

## Potential Lessons from the Truly Departed

Let us reiterate: Our simple inquiry was not designed in order to construct significant generalizations from a large or unique database. Rather, we hoped to openly and honestly connect with those volunteer families who might later serve as informants for a study of home educators' curriculum and instruction practices. Towards this ultimate end, we posed two simple questions could might permit us to discover certain angles and issues related to home education which might not yet have been developed within this growing body of scholarship, and permit our respondents to remain anonymous or self-identify as a statement of further interest.

While public schools in Pennsylvania and across the United States seem grudgingly headed toward positions of greater interactive support for home educators, they do so, in part, to recoup moneys lost when "home" students do not appear on public school roles. Beyond this mercenary motivation, reconciliation is sought in the name of accountability and control. Maralee Mayberry believes, for example, that "a significant proportion" of home educators who are permitted to have a say in how new relationships get negotiated between themselves and their local public schools will, over time, "accept some guidance and standards from states and public schools" (Cohen, 1995, p.6). Meanwhile, few efforts are made to critically reflect upon what home-based educators have to say "about learning, about educational policy, and about the strength and viability of the institution of schooling" (Van Galen & Pitman, 1991a, p. 5).

We believe that our preliminary inquiry, when seen in light of the existing knowledge about home-based teachers and learners, contains several important inferences of value to those engaged in school reform efforts. To begin, don't oversimplify people and their concerns. Public school curricula remain "godless" in the eyes of primarily religious-motivated home educators (Van Galen's Ideologues). And though issues around the "wall of separation" between the secular and spiritual aspects of public schooling in this country continue to proliferate in all venues of public discourse, our data suggest that such issues are typically interwoven with others having to do with social and pedagogical values. Complex issues like these provide openings where people can explore and attempt to untangle their concerns in an effort to communicate their differences and seek commonalities.

The greatest area of concern registered by the home educators represented here pertains to parents' dissatisfaction with schools in which their children could not learn and grow strong in appropriate ways (Van Galen's Pedagogues). Rather than place their children within environments they characterized as too quick to produce and act according to labels (e.g., behavior problem or slow learner), or too academically challenging or unchallenging, most of these families claim to have given up on the possibility of that ever happening. For these families to dismiss those opportunities which can perhaps best be provided through the educational agency of school is a tragic loss which affects everyone who cares about civic America.

The most complicated and pertinent message about the state of public school affairs we find within our data pertains to home educators' concerns about "peer influence" -- a message all but lost when oversimplifying the Ideologue/Pedagogue categories. Variously referred to as concerns about the effects of urbanization and modernization (Mayberry & Knowles, 1989) or the quality of socialization (Mayberry, 1989), parents of all religious, ideological, and social persuasions in our sample are removing their children from U.S. public schools on the basis of "peer concerns" (for additional support for and elaboration of this position, see Aiex, 1994; Gladin, 1987; Knowles and Others, 1994; Morgan & Rodriguez, 1988; Pike, 1992). The message here is that schools are simultaneously feeding and reflecting broader social and cultural changes which are considered inappropriate by growing numbers of people.

This critique of schools is not new. The 26th annual Phi Delta Kappa/Gallop Poll of attitudes toward public schools indicates that among the top four problems faced by schools and communities are "fighting/gangs/violence," "lack of discipline," and "drug abuse" (Elam, Rose, & Gallop, 1994). Indeed, concerns about discipline and drugs have been uppermost in the minds of respondents over the past 25 years of such polls (Elam, Rose, and Gallop, 1993).

And while poll respondents carefully complete these Gallop surveys, Pennsylvania's home educators continue in growing numbers to remove their children from socially and culturally complicated public school environments. In our state, the number of school-aged children educated at home doubled between 1990 and 1992 as the number of home education support groups climbed to more than 100 (Richman, 1994).

That our sample of home educators comes from a largely rural Pennsylvania community underscores the need for concerned school reformers to confront the porous nature of the school/community inter-relationship head on -- not in an attempt to more successfully isolate its school inhabitants, but rather in an effort to identify and better understand larger problems, construct and critique desirable alternative visions, and determine appropriate collective actions (Note 3). Such opportunities provide a site where parents, educators and community members struggle through their distinct and reinforcing roles and responsibilities -- a site where the realization that various educational agencies must jointly participate in the transmission of cultures to our youth cannot be ignored.

## Conclusion

With so many public school educators diligently at work to bring renegade parent educators back in line in terms of the products of public schooling (test scores, content coverage, minutes on-task, etc.), we believe that those committed to public school reform ought to pay a different sort of attention to them.

Confronting a changing culture is the order of the day for a public school machine slowly becoming obsolete within an increasingly conservative, libertarian effort to ignore an inevitably postmodern world (see Doyle, 1992). In this world, absolutes are fading, demands upon schools have increased to the point where individual learning and development can no longer be taken for granted, and balkanization, fear and ennui have overwhelmed civic-mindedness. And while schools have obvious and crucial educational and cultural responsibilities in light of this world, they are not alone.

To address these issues, Jane Roland Martin urges schools to return to an earlier position wherein they shared their responsibilities with other educational agents -- particularly with the home. This change will require that those who represent schools see themselves, again, as members of "the whole range of cultural custodians" and accept that "school has much to gain from treating other educational agents as partners rather than as humble assistants or else dangerous rivals" (10). Doing so also creates the need for all educational agents to understand, appreciate, and accept responsibility (and thus, be accountable) for the cultural work at hand. In her words: if we can envision an array of institutions, all of which share the tasks of preserving our vast cultural assets, see themselves and are seen by others as legitimate educational agents, and work together to transmit the [cultural] wealth, we will at least have a better idea of what to strive for. (1996, p. 10)

We choose to see home educators as thoughtful and important critics of public schooling who have decided to assume their responsibilities as what Henry Giroux terms "cultural workers" at great personal cost and uncertainty. Parents who educate their children at home do so at considerable cost (Bishop, 1991; Reynolds & Williams, 1985; Williams and Others, 1984). It is "an arduous option" (Lines, 1983, p.183) to educate one's children at home; as Virginia Seuffert (1990), a home-teaching mother notes, "Home-schooling dominates your time and demands a certain energy level that not everyone has" (p. 74).

Nonetheless, the number of home educators continues to increase nationwide -- a fact that should put everyone committed to the ongoing reformation of public schools on notice. That so many families we contacted in rural Pennsylvania have exited the public schools solely or primarily for "pedagogical" reasons, that more than one third remove their children because of "peer influence" concerns, and that so few parent-teachers can imagine their children returning to those exited public institutions ought to tell us something not only about our neighbors but about ourselves. Perhaps it's time for us to consider the possibility that these "truly departed" represent important voices in our continuing efforts to reform schools in light of our changing world.

## References

Aiex, N.K. (1994). Home schooling and socialization of children. ERIC Document Reproduction Service No. ED 372 460.

Bishop, C. (1991). Home schooling parent support groups in Kansas: A naturalistic inquiry into their concerns and functions. Ph.D Dissertation, University of Kansas. ERIC Document Reproduction Service No. ED 338 368.

Calvery, R., and others. (1992). The difference in achievement between home schooled and public schooled students for grades four, seven, and ten in Arkansas. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN, November 11-13. ERIC Document Reproduction Service No. ED 354 248.

Clark, C. (1994). Home schooling. *C Q Researcher*, 4 (33), 769-792.

Cohen, P. (1995). Schooling away from school. *Education Update*, 37(6), 1,6,8.

Doyle, D.P. (1992). The challenge, the opportunity. *Phi Delta Kappan*, 73(7), 512-520.

Elam, S.M., Rose, L.C., & Gallop, A.M. (1994). The 26th annual Phi Delta Kappa/Gallop poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 76(1), 45-56.

Elam, S.M., Rose, L.C., & Gallop, A.M. (1993). The 25th annual Phi Delta Kappa/Gallop poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 75(2), 137-151.

Frost, G. (1988). The academic success of students in home-schooling. *Illinois School Research and Development*, 24(3), 1111-117.

Gladin, E.W. (1987). Home education: Characteristics of its families and schools. Ed.D. Dissertation, Bob Jones University. ERIC Document Reproduction Service No. ED 301 925.

Groover, S.V., & Endsley, R.C. (1988). Family environment and attitudes of homeschoolers and non-homeschoolers. ERIC Document Reproduction Service No. ED 323 027.

Harris, J.J., & Fields, R.E. (1982). Outlaw generation: A legal analysis of the home-instruction movement. *Educational Horizons,* 61(1), 26-31, 50-51.

Jeub, C. (1994). Why parents choose home schooling. *Educational Leadership,* 52, 50-52.

Klicka, C.J. (1990). Home schooling in the United States: A statutory analysis. Paeonian Springs, VA: Home School Legal Defense Association.

Knowles, J.G., & Others. (1994). Home education as an alternative to institutional education. *Educational Forum,* 58(3), 238-243.

Lines, P. (1991). Home instruction: The size and growth of the movement. In J. Van Galen & M.A. Pitman (Eds.), *Home schooling: Political, historical, and pedagogical perspectives* (pp. 9-42).. Norwood, NJ: Ablex.

Lines, P.M. (1983). Private education alternatives and state regulation. *Journal of Law and Education,* 12(2), 189-234.

Litcher, J.H., & Schmidt, S.J. (1991). Social studies in the home schools. *Social Education,* 55(4), 239-241, 248.

Mahan, B.M., & Ware, B.J. (1987). Home schooling: Reasons some parents choose this alternative form of education, and a study of the attitudes of home schooling parents and public school superintendents toward the benefits of home schooling. ERIC Document Reproduction Service No. ED 286 624.

Marshall, J.D., & Sears, J.T. (1985). John Holt: In memory. *Changing Schools,* 13, 5.

Martin, J.R. (1996). There's too much to teach: Cultural wealth in an age of scarcity. *Educational Researcher,* 25 (2), 4-10.

Mayberry, M. (1989). Home-based education in the United States: Demographics, motivations and educational implications. *Educational Review,* 41(2), 171-180.

Mayberry, M., & Knowles, J.G. (1989). Family unity objectives of parents who teach their children: Ideological and pedagogical orientations to home schooling. *Urban Review,* 21 (4), 209-225.

Meighan, R. (1984). Political consciousness and home-based education. *Educational Review,* 36(2), 65-73.

Morgan, P.T., & Rodriguez, R.C. (1988). Home school parents: A rural survey. *Rural Education,* 10(1), 15-18.

Pike, B. (1992). Why I teach my children at home. *Phi Delta Kappan,* 73(7), 564-565.

Ray, B.D. (1988). Home schools: A synthesis of research on characteristics and learner outcomes. *Education and Urban Society,* 21(1), 16-31.

Ray, B.D., & Wartes, J. (1991). The academic achievement and affective development of home-schooled children. In J. Van Galen & M.A. Pitman (Eds.), *Home schooling: Political, historical, and pedagogical perspectives* (pp. 43-62 ).. Norwood, NJ: Ablex.

Reynolds, P.L., & Williams, D.D. (1985). The daily operations of a home school family: A case study. Paper presented at the Annual Meeting of the American Educational Research Association, March 31 - April 4. ERIC Document Reproduction Service No. ED 256 080.

Richman, H.B. (1994). Home schooling: The oldest educational invention. In C.E. Greenwalt II (Ed.), *Educational innovations: An Agenda to frame the future* (pp. 221-239). Lanham, MD: University Press.

Seuffert, V. (1990). Home remedy: A mom's prescription for ailing schools. *Policy Review*, 52, 70-75.

Stough, L. (1992). Social and emotional status of home schooled children and conventionally schooled children in West Virginia. Masters Thesis, University of West Virginia. ERIC Document Reproduction Service No. ED 353 079.

Tipton, M. (1990). An analysis of home-schooled children's Comprehensive Test of Basic Skills results and demographic characteristics of their families. Masters Thesis, Antioch University. ERIC Document Reproduction Service No. ED 336 208.

Van Galen, J.A. (1988). Ideology, curriculum, and pedagogy in home education. *Education and Urban Society*, 21(1), 52-68.

Van Galen, J. (1991). Ideologues and pedagogues: Parents who teach their children at home. In J. Van Galen & M.A. Pitman (Eds.), *Home schooling: Political, historical, and pedagogical perspectives* (pp. 63-76).. Norwood, NJ: Ablex.

Van Galen, J., & Pitman, M.A. (1991a). Introduction. In J. Van Galen & M.A. Pitman (Eds.), *Home schooling: Political, historical, and pedagogical perspectives* (pp. 1-5 ). Norwood, NJ: Ablex.

Webb, J. (1989). The outcomes of home-based education: Employment and other issues. *Educational Review*, 41(2), 121-133.

Williams, D.D., & Others. (1984). Understanding home education: Case studies of home schools. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 23-27. ERIC Document Reproduction Service No. ED 244 392.

## Notes

1. We wish to acknowledge and thank Gary Knowles and Pat Shannon for their helpful and insightful conversations with us as we worked to write and revise this piece.
2. Given the distinction between the general terms "education" and "schooling," wherein the latter is typically associated with bureaucratized and impersonalized institutional arrangements designed to promote the former, we have chosen to employ the term "home education" for our work here.
3. Dr. Betty Beach explores rural home educators' situations in particular. She can be reached via e-mail at bbeach@maine.maine.edu for specific information and dialogue.

## About the Authors

J. Dan Marshall
Associate Professor
146 Chambers
University Park, PA 16802
(814) 865-6569
FAX: (814) 863-7602

James P. Valle
Elementary Teacher & Home School Evaluator
Donegal School District

Mount Joy, PA 17552

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@pisces.rutgers.edu*
John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Sherman Dorn
*sfxj9x@scfn.thpl.lib.fl.us*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 13** **August 15, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS
ARCHIVES.Permission is hereby granted to copy any
article provided that EDUCATION POLICY ANALYSIS
ARCHIVES is credited and copies are not sold.

---

# Implementing AIDS Education: Policies and Practices

**Grace C. Huerta**
**Utah State University**

*graceh@cc.usu.edu*

**Abstract:**
The world has been challenged by the AIDS epidemic for 15 years. In 1985, the U.S. Department of Health and Human Services, Centers for Disease Control, allocated funds to all state departments of education to assist schools in the development of AIDS education policies and programs. Yet, these policies do not ensure that all students receive effective AIDS education. On September 21, 1991, the Arizona Legislature passed Senate Bill 1396, which requires public schools to annually provide AIDS education in grades K-12. The bill was rescinded in 1995. With prohibitive curriculum guidelines, limited teacher training opportunities and tremendous instructional demands, this educational policy was implemented in disparate forms. By examining the perspectives of the Arizona educators (representing three school districts), this qualitative study reveals how teachers ultimately controlled the delivery and nature of AIDS instruction based upon personal values, views of teacher roles, and their interpretation of the mandate itself.

## INTRODUCTION

Adolescents are particularly vulnerable to contracting the Human Immunodeficiency Virus, the virus considered by many to be responsible for the opportunistic infections associated with AIDS. Because of the disease's latency period, more than 20 percent of persons reported with AIDS in the United States are under the age of 30 and were probablyinfected during their teens (WHO, 1993). Whether out of curiosity, the wish to experiment, peer pressure or low self-esteem, teenagers engage in pre-marital, high risk sexual intercourse. Many activists maintain that without a vaccine, the only means of AIDS prevention is through education (Aiken, 1987).

In 1985, the U.S. Department of Health and Human Services allocated block grants to state departments of education to assist schools in the development of AIDS educational policies, teacher training programs and curricula. However, the development of specific AIDS curriculum guidelines has been particularly challenging. Politicians and educators are cognizant not only of resource allocation but of constituent and public opinion associated with the stigma surrounding AIDS. As a result, the task of delivering a curriculum which responds to such questions as: What is AIDS?; How is it transmitted?; How can it be prevented? and; Who should teach this? remains politically charged.

In addition, few teachers have the training and theoretical tools to address such questions in the classroom (Aiken, 1987; Dodds, Volker & Viviand, 1989; Eckland, 1989; DiClementi, 1990; GAO, 1990; Nadel, 1990; SIECUS, 1991; NASBE, 1993; Popham, 1993b). Therefore, the intent of this research is to reveal how the teachers of Arizona, interpreted, and implemented a mandated AIDS education policy despite these challenges.

## DATA COLLECTION

The flexibility of qualitative methodology allows for the use of multiple methods and strategies for analysis. Qualitative methods call for thick, rich descriptions of processes and are concerned with the meanings which participants attribute to social interactions and situations (Geertz, 1973). This inquiry is based upon participant behaviors, actions and meanings, not assumptions about these constructs.

It is for these reasons that a qualitative approach was employed for this study of the "practice" of AIDS education. The following data collection techniques were used: direct observations, participant observation, structured interviews of participants and the analysis of documents.

### Direct Observations

The observation of state district AIDS teacher training sessions, enabled the researcher to witness a variety of interactions, activities and responses regarding the implementation the Arizona AIDS K-12 education mandate. Because of the unique social construction of AIDS, the observation of participant interactions helped establish beliefs regarding the mandate during different stages and levels of its implementation. Observation sites included the following teacher training sites: the Arizona Department of Education (ADE) Comprehensive Health Department; district A-suburban (K-8) health education unit; district B-urban (K-12) comprehensive health education unit; district C-rural (9-12) health education unit; elementary and secondary teacher training sessions held at District A,B,C designated school sites; three parent/community AIDS education information meetings, and; three school board meetings.

Attending state-sponsored teacher training sessions helped to establish contacts with teachers who have served or would serve as site AIDS instructors. These contacts also helped to identify the nature of various district and school AIDS education efforts. These observations began in October 1992 and continued through December 1993. Observations of parent meetings, ADE curriculum development and organizational activities took place between November 1992 and June 1993.

Three districts, representing rural, suburban and urban settings were randomly selected from Maricopa County, the largest county in the state.

Requests to observe classroom AIDS lessons were, for the most part, denied, as principals (five from each of the three districts) refused to allow access to view AIDS instruction. The principals gave a variety of responses, ranging from an unequivocal "No," (with little explanation attached), to directing me to review district policy (because that is what they felt their teachers taught), to assuring me their schools had already delivered the ADE curriculum in 1992-'93 and they had not yet established plans for the 1993-'94 academic year.

### Interviews

Structured interviews revealed the multiple perspectives and views of those charged with implementation of the AIDS education mandate: policy-makers, state and district administrators, as well as teachers. Through the course of these structured interviews, questions which pertain to beliefs regarding Arizona AIDS education policies, resource allocation, curriculum development, teacher training and instructional practice were raised.

The interviewees included members of the Governor's Task Force on AIDS; the Arizona School Boards Association, and; the Arizona Legislature Education Committee;

Persons charged with the implementation of the mandate were also interviewed. Taking part were the Arizona Department of Education AIDS specialists, three district health education administrators; ten principals; nine secondary public school teachers representing three districts, and; nine elementary public school teachers representing three districts.

The perspectives of these participants helped determine relevant background information and context, as well as identified the antecedents which prompted the creation of the policy in its present form. The tape-recorded interviews (see Appendix A interview protocol) took place at the state department of education, district teacher training sites, and specific school sites.

Data was then transcribed and coded for the purposes of analysis. Core categories that emerged include policy development issues, teacher training processes, barriers to instructional practice, compliance circumvention, and evaluation methods.

### The Use Of Documents

Numerous documents were collected and analyzed. They played a vital role in providing information about organizational structure, funding and evaluation efforts. The following documents were reviewed: the Arizona AIDS education mandate, S.B. 1396; legislative minutes related to S.B. 1396; CDC Guidelines for Effective School Health Education To Prevent the Spread of AIDS; an external evaluation of the Arizona Department of Education HIV/AIDS Prevention education program; Arizona Department of Education K-12 HIV/AIDS curriculum guides; parent information and consent forms; staff training materials; the Council of Chief State School Officers' Profile of State AIDS Education Survey Results, and; the National Association of State Boards of Education report on AIDS and School Health Education--State Policies and Programs.

The analysis of documents provided information regarding the CDC guidelines pertaining to funding, records concerning legislative sessions regarding policy development, the Arizona AIDS mandate itself, curriculum standards and evaluation concerns. This content analysis yielded information about the organizational norms derived from participant beliefs regarding AIDS education efforts. Since the review of documents is an unobtrusive research method, this began a particularly important part of this study, since the sensitive nature of the issues involved caused the participants to be hesitant to respond freely to the interview questions. Also, the review of documents helped generate additional interview questions which were otherwise overlooked.

Documents which reflected quantitative data collection were also examined. Since different kinds of research questions can be addressed when using multi- methods, quantitative data provided, for example, such data as: the percentage of schools providing AIDS instruction by school administrators; the percentage of the schools providing AIDS instruction as reported by teachers; the distribution of AIDS education provided at various grade levels, and; the percentage of obstacles to AIDS instruction as perceived by the teacher themselves. Together, each quantitative and qualitative methods help to stimulate research question as well as establish assertions regarding the findings. In addition, by combining qualitative and quantitative methods, bias was reduced.

This interpretive research sought to address questions related to the "hows and whys" of the implementation process The final phase of this study began with the synthesizing of the data for the purposes of identifying categories of phenomena and the relationships between them. This required a careful review of tape recordings, transcripts, documentation and fieldnotes. From these categories, ideas and themes were generated. Listening to personal accounts shared by the participants also helped to establish rapport, and raise addition research questions.

By identifying and categorizing problematic areas, a focused synthesis followed. A "focused synthesis" (Doty, 1982) consists of the selective review of information, relevant to the policy study's research questions.

In general, this policy study employed naturalistic methods in order to arrive at a valid, corroborated, interpretation of the data required to answer the research questions. These methods included: direct observations, one-on-one structured interviews, analysis and deconstruction of documents, the review of quantitative data, the creation of core categories and themes, and the focused synthesis of the data set. The findings were then integrated into an overview of the development and implementation of the AIDS education policy by its practitioners--professional educators.

## AIDS INSTRUCTION IN ARIZONA-- DISTRICT A SCHOOLS

Society is ambivalent about the role for teachers when entrusting them with a set of life choices and values to put before their students. It is the educator, however, who is charged with implementing mandates and devising programs whose foundations rest on moral questions. The following cases illustrate how Arizona teachers strived to implement AIDS education programs in both secondary and elementary public school settings.

In September 1991, the legislature of the state of Arizona passed its own AIDS education mandate, Senate Bill 1396. It requires Arizona public schools to provide AIDS education in grades K-12 annually. Each district is free to develop its own course of study for each grade. According to the mandate, the curriculum must reflect the following: 1) grade level appropriateness; 2) medical accuracy; 3) abstinence; 4) drug prevention, and; 5) modes of AIDS transmission. In addition, the curriculum cannot promote a homosexual lifestyle, portray homosexuality as a positive alternative life-style or suggest that some methods of sex "are safe forms of homosexual sex" (ADE, 1992, 2). All school districts are required to hold parent meetings to describe the curriculum prior to providing AIDS instruction. In addition, each school must notify parents of their right to withdraw their children or "opt out" of AIDS instruction if they so choose.

In 1992, District A (grades K-8) developed a program based upon the state's AIDS curriculum recommendations. The district committee also sought to adapt the state curriculum to meet specific areas of concern. The committee, comprised mostly of District A nurses, felt the AIDS curriculum should emphasize the disease process and in particular, should include a discussion regarding common illnesses and how they are contracted. This focus resulted in the development of supplemental lesson plans for teachers to use in conjunction with the state curriculum. In addition to the state department of education-sponsored AIDS in- services, District A offered its own training sessions for their school personnel.

All District A teachers are responsible for delivering the AIDS curriculum. Because of the nature of the self-contained elementary school classroom, the teachers are instructed through the district and state teacher training in-services that the AIDS lessons should be presented within the context of the health education unit. The teachers of District A appeared to be comfortable when providing the curriculum within this unit, that is, if they had participated in an AIDS teacher training in-service. Two teachers noted:

> At first I thought there was no way I could teach this curriculum. To be honest, what I knew about AIDS I learned from TV, the papers. After I went to the trainings, I felt better equipped to teach the subject. But that doesn't mean I felt more comfortable about it. So I taught a few of the lesson plans during health ed, when I talked about the body, the immune system, and how it protects us. I used the lessons that were appropriate for my class (District A teacher 1, interview, January 22, 1994).

> In my kindergarten class, I talked to the students about germs, infection and how they can lead to such illnesses as colds. We also discussed hygiene. That's what the AIDS lesson plans cover at this level, that's what our curriculum focuses on. It's odd this basic health information falls under the AIDS ed umbrella. But I think the kids will come to understand the illness if their future teachers give the lessons in the correct sequence. I felt comfortable fitting it into my health ed unit, and the trainings helped me to understand how to do that (District A teacher 2, interview, January 23, 1994).

Other teachers had not yet participated in a state department of education or district AIDS teacher training in-service. Although they were expected to begin providing AIDS instruction in the 1992-'93 academic year, there were District A teachers who did not take part in the trainings, nor did they deliver the curriculum to their classes. None of these teachers described any punitive measures taken against them by their site or district administrators. One fifth grade teacher explained:

> I didn't attend an AIDS in-service last year. I probably shouldn't be telling you this. But it's not because I didn't want to teach it. It's because I just didn't have the time. I teach third grade LEP (Limited English Proficient) and I spend a lot of time preparing for my class. The state only offered a limited number of spaces for their trainings and when I signed up, the in-services were full. The district in-services are after school, and I'd rather spend that time with my own kids. So I didn't teach the AIDS curriculum last year [1992-'93]. But I did go to the department of ed training this fall. Then I found out the curriculum wasn't translated into Spanish. I'll have to do that myself, so I probably won't give the lessons until sometime this spring [1994], (District A teacher 3, interview, January 29, 1994).

> Every year it seems like we have to teach something else. So I wasn't thrilled about this AIDS education mandate. Last year (1992- '93) I didn't go to an in-service and I didn't teach the curriculum because I was busy and wasn't prepared. I received a memo from my principal that I had to attend one this year. Why should I? Am I getting paid extra? I'll probably teach the lessons this semester [1994], but I still feel this is just another thing we have to do. This [AIDS education] should be taught at home (District A teacher 5, interview, February 8, 1994)).

For the teachers who did present the AIDS curriculum in the classroom, they found the controversial nature of the content to be tempered by the health concepts introduced at each grade level. Initially, there were concerns for those students' whose conservative parents would choose to opt their child out of the AIDS education program. According to the District A Health Education Administrator, only one student reportedly "opted-out" of the program last year. She noted,"Most parents want us to teach their children about the disease and how not to get it" (interview, November 16, 1993).

The District A AIDS teachers acknowledged that once they felt familiar with the curriculum, they were able to integrate it within their health education or family health units. One sixth grade teacher recalled:

> When I began the AIDS curriculum, I was worried about how it would work with the rest of the health unit. At this grade level we begin to talk about some of the specifics of the disease, such as the transmission of AIDS through bodily fluids. We did an activity that I learned about at the 2-day training. We mixed food coloring, water and chemicals in one another's test glasses and shared 'fluids.' If the mixture changed colors, we learned whether we were 'infected' or if we infected someone else. The students enjoyed it. During our discussion, I introduced the topic of abstinence. I stressed to my students that abstinence is the best way to minimize the risk of getting infected (District A teacher 6, interview, February 7, 1994).

One District A kindergarten teacher, found her students and their parents to be concerned about a variety of issues:

> During our parent meeting, the principal had asked a teacher from each grade level to present an overview of the curriculum. We wanted to assure the parents what was being taught was grade appropriate. I discussed the kindergarten section, answered their questions, got some feedback. I learned about some of their concerns, which were mostly about whether there was going to be a discussion about sex and homosexuality. They seemed relieved when I told them we would be discussing things like germs, infection and hygiene, not sex.

The biggest challenge faced by the District A AIDS teachers in the classroom pertained to the students' concerns about where the virus came from, who it affects and how it is transmitted. Following the introduction of basic health concepts, the discussion of the immune system, and the definition of AIDS and AIDS, the teachers found it difficult to dispel the myths about the virus, as required by the mandate. The District A AIDS teachers were continually faced with a variety of misconceptions held by students, regardless of grade level. The teachers acknowledged that while the AIDS in-services prepared them to discuss basic issues, student concerns were more specific and frank. The teachers found it difficult to anticipate all of the students' questions and provide responses which did not violate the mandate's guidelines.

District A AIDS teachers' discussions with students often reflected the restrictions set forth by the state curriculum guidelines. The guidelines limited discussions about how AIDS is transmitted, whether the acts of sexual expression be heterosexual and homosexual. Instead, students received information about where AIDS could be found (i.e. blood, semen, and vaginal fluids). Students were informed that only by maintaining an exclusive monogamous relationship with an uninfected partner would they be insured of eliminating the risk of contracting AIDS. While this promotes the mandate's abstinence message, the teachers did not address the issues regarding the identification of an infected partner, nor did it allow for discussion about why students should remain monogamous, only that they should. District A teachers tended to equate sexual expression with punishment, disease, and eventually, terminal illness:

> I told my eighth grade girls that they should wait until marriage before having sex. That way they could get to know their partner. I told them that pre-marital sex leads to trouble, teen pregnancy, sexually transmitted diseases, and AIDS. I told them, 'Do you want to live or die?' These girls need to learn to say 'No' (District A teacher 8, interview, February 21, 1994).

> These kids need to know that homosexuality kills. That's not a politically correct thing to say, I know. But the majority of people who are contracting the disease are still homosexuals. When I taught the seventh grade boys last year, I told them that this behavior is not okay. I don't care how many gay rights bills are passed and how many of them march on Washington. Homosexuality is wrong and gays are dying because of it. That's all really what these kids need to know (District A teacher 9, interview, February 22, 1994).

In the 1992-'93 academic year, various District A teachers presented the AIDS curriculum in the manner they felt was in compliance with the 1991 mandate. Although the teachers presented the curriculum within the context of the health education unit, none introduced material beyond what was required of them. Follow-up activities generally consisted of post-tests or quizzes which sought to measure student knowledge about the content presented to them regarding such topics as: personal hygiene, the immune system, the definition of AIDS and AIDS and the importance of abstinence.

## AIDS INSTRUCTION IN DISTRICT B SCHOOLS

Prior to the passage of the Arizona AIDS K-12 education mandate in September 1991, AIDS education was not a formal part of the District B curriculum. Following the passage of the bill, District B set out to develop its own curriculum (although very similar to that of the state department of education). District B made a concerted effort to provide training for teachers in order to emphasize the recommendations established by its own Health Education curriculum committee. Similar to the state recommendations for providing AIDS instruction, District B strives to integrate the AIDS content into a school's pre- existing health education program. Unique to District B is its development of a separate health and safety unit, the "Community Survival Curriculum," for those parents who have opted their children out of the AIDS and sex education programs.

Occasionally, District B principals would acknowledge that not all of the teachers were willing participants in the AIDS teacher training in-services. When principals would send a memo to a teacher who had not attended an in-service, they would, at times, encounter resistance from those teachers. One principal recalled:

> I had teachers who didn't want to attend any of the in-services. And I can't say that I blamed them. These teachers are overwhelmed by all of the things they have to do in the classroom. They aren't interested in teaching a curriculum they had very little input in developing. They have enough on their minds, lesson plans, test preparations, motivating kids. Of course I tell them they are required to attend the in-services, but I let them know that I empathize with them and that I understand their reluctance (District B principal 4, October 19, 1993).

Of the District B teachers who did not attend an AIDS teacher training in-service, they maintained it was not because of ideological or moral concerns. They reiterated that while they were aware of the seriousness of the AIDS epidemic, they had other concerns which they felt were more pressing:

> I knew I was supposed to go to one of those in-services, but I never did. I had too many other things to do. Teaching third grade keeps me busy. And besides, I've seen the [AIDS] curriculum and it basically just covers health issues. And that's what I teach in my class anyway, so I don't see why it is so important that I go to one of those things. I'll probably go to one this year, at the district, at least those are shorter than the state trainings (District B teacher 2, interview, October 25, 1993).

Of the District B teachers who did attend an AIDS in- service, they found that by presenting the curriculum within a broad health education framework enabled them to discuss such issues as disease transmission, the immune system and hygiene. However, some teachers found that this approach also revealed the limitations of the AIDS curriculum. One eighth grade teacher said:

> Even with the [AIDS] training I had, I was concerned about how I was going to present this topic to the girls. You know, at this level we separate the kids by gender, as we get into the Family Life [sex education] unit. The girls need to know about how their bodies function and how to have to take care of themselves in a variety of ways, from nutrition to exercise, to decision-making. We need to go beyond discussions about menstruation and reproduction, and actually, the AIDS unit starts to get us there. When you really sit and think about it, this material is very difficult to deliver. I want to talk about health as a whole, but how can I do that when I can't even talk about the issues that matter to them. I tell the girls that AIDS is carried in blood, semen, and they have a lot of questions, but I know my answers can't stray away from the emphasis on abstinence. I have to stick to the biological information (District B, teacher 4, November 1, 1993).
>
> These kids [seventh graders] say more than I can teach about AIDS. Well, at least they're not afraid to speak their minds. They know it's spread through sexual intercourse but they're not exactly sure how. Of course I'm not supposed to mention those kinds of things. But they've heard about Magic Johnson. And I have students who say 'Only faggots get that.' They have 'sound bites' of information, some are true, some are false. So in the long run, it's ridiculous what we can and cannot talk about. How can I get at these myths when I can't talk about them? (District B teacher 5, November 2, 1994).

When confronted with questions which were not directly addressed in the state or District B AIDS curricula, the teachers often had to separate what they felt was practical information from the content they were required to deliver. faced with conflicting kinds of information posed a special challenge to the District B instructors who did not want to leave misinformation unaddressed. And yet, they did not want to delve into topics which were "off limits" according to the state curriculum guidelines. For the most part, these topics, and the myths surrounding them, concerned the ramifications of homosexuality, monogamy and promiscuity. The District B teachers were aware that these issues were taboo, even though they were the same topics the upper grade students were most interested in. One eighth grade teacher recalled:

> Actually, these girls love the idea of monogamy. It's so romantic to them. They don't see themselves getting AIDS. I brought in an news article. It was about a woman in her 30s, with three kids. She was AIDS positive, and she got it from her ex-husband. My students were intrigued by the fact that the ex-husband 'did it to her,' and they were glad to learn that he had died. 'He deserved it,' was their response to that story. They felt 'The woman didn't deserve it. She was only with her husband and look what happenedto her.' The girls seemed genuinely troubled by the idea that men fool around, as if that's an accepted part of their nature. 'That's why you really have to be a good wife,' one student said. Now that really stuck in my mind, because then we got into a discussion about what makes a good wife. 'Pleasing your man,' was the general response. One thing about teaching the AIDS curriculum, it really opens up a can of worms. But I'm really not so sure that the abstinence message is just a retelling of the fairy tale (District B teacher 7, interview, February 11, 1994).

The Arizona AIDS curriculum did have its supporters. Several teachers expressed that they were very comfortable presenting the lesson plans within a biological context. They did not feel limited by the mandate's focus on abstinence, or the exclusion of the homosexuality as a topic of discussion. In some ways, the teachers felt that those guidelines made the delivery of the mandate much less controversial and consistent with their personal values. The teachers recognized that the guidelines also made the curriculum more palatable for parents. One sixth grade teacher said:

> When I taught about AIDS, I told my students [sixth graders], right off the bat, we are not going to talk about homosexuality. We are not going to talk about how to have sex outside of marriage. We're not allowed to, and I don't want to because I don't believe in either of those lifestyles. I told them we were going to define AIDS, AIDS and how this disease infects the immune system. Once I said that, the kids knew I wasn't playing games and they shouldn't play games with their lives (District B teacher 9, interview February 23, 1994).

## AIDS INSTRUCTION IN DISTRICT C SCHOOLS

In August, 1991, while the passage of the Arizona AIDS K-12 mandate was still one month away, District C high schools had already organized an AIDS education committee (composed of volunteer educators) and had developed a curriculum which would generally mirror that which was published by the state in 1992. By the 1992-'93 academic year, committee had designed an "AIDS Awareness Week" to present to staff and students district-wide.

Teachers responsible for providing AIDS instruction in District C schools did so on a voluntarily basis, regardless of content area specialization. "Teacher teams" would travel from class to class and provide the AIDS instruction throughout the week. Once District C AIDS classroom instruction began, teachers were at times taken aback by their students' responses. One AIDS instructor found one student's notions about the disease to be extremely troubling. She explained:

> We are required to emphasize to our students abstinence and to make better choices in risky situations. But consider the kind of information students have. One student talked to me after a presentation. She said she felt she was a virgin because she had anal sex. That was her form of birth control....'That way I won't get pregnant' she said. But what she didn't really understand was that anal sex is the highest AIDS risk behavior! (District C AIDS teacher 4, interview, December 14, 1993).

The AIDS instructors faced a number of questions from students which were frank and were not addressed in the state curriculum. Instructors found eleventh and twelfth grade students to be interested in whether or not they could contract AIDS by 'French kissing,' having intercourse during menstruation or while engaging in oral sex. Such questions were easily recalled by the AIDS instructors because they were sensitive in nature, miles away from the issue of abstinence and were the most difficult to answer. One teacher said:

> These kids want to know more than abstinence. They demand it. When the students ask these kinds of questions, there's always a bit of laughter. Part of it is because they want to see how I'm going to respond. I try to approach these questions from the biological angle. That may seem like I'm avoiding the issues. Then I tell them AIDS istransmitted through bodily fluids and which includes blood and semen. I tell them there are still many questions that researchers don't have the answers yet. My bottom line is they're not going to get AIDS if they sit on a toilet. They're not going to get it if someone spits on them. Or if a gay person sits next to them on the bus. The bottom line here is that these kids want to know the answers, and these kids are the juniors, the seniors, the ones who about to begin their adult lives. (District C AIDS teacher 3, interview, December 13, 1993).

Having "the right answers" proved to be a chalienge to the site AIDS instructors. Although many participated in the ADE sponsored teacher training in-services, their responses to student questions did not always come quickly or easily. While several of the presenters had some form of health education background, this did not necessarily mean they were able to address all of the questions the students raised. Group activities helped ease the pressure some teachers experienced while trying to answer sensitive questions:

> At the beginning of our third session, after the biology and the discussion about transmission, I sensed that the students still felt this was a gay disease and the dialogue just shut down. They didn't participate in the discussions and I felt that I was just talking to myself. So then I told them, 'Let's play, 'What's My Line?' This is a role play in which the kids receive cards identifying different types of people, for example, a single mother, a male dancer, a father of five. It's up to the other students to determine who is AIDS positive. They ask the ones who received the cards questions. The cards they hold have scripted responses written on the back, but the kids are free to elaborate. So of course, the class expects the AIDS infected person to be the male dancer. But they guess wrong, it's the father of five. So the class is confronted by their own stereotypes. In their silence afterward, I felt that they got the message (District C AIDS teacher 5, interview, October 20, 1993).

This is not to say that all students were responsive to the instruction. A number of the District C AIDS teachers found it difficult to engage students in discussions and cooperative learning activities. While it may have been more efficient to distribute the AIDS teaching resources to those classes who requested a team, this process also had its drawbacks. Establishing rapport with a new class, and with a sensitive curriculum to deliver, was not always the most effective context for instruction. The "AIDS teaching teams" found that students were not always willing to share their ideas, or were uncomfortable discussing sensitive topics in front of their peers. Often times it became difficult for the teams to gauge whether or not their instruction was, "Sinking in," as one teacher wondered.

Because instruction was not provided in a broad, comprehensive health context, some AIDS teachers felt awkward when beginning instruction. Classes were not prepped before the AIDS instruction began. The teams entered different classroom settings and began discussing such issues as: the immune system, sexually transmitted diseases, human sexuality, stereotypes and decision-making techniques:

> Going into a new class to teach about AIDS was something I took too lightly. Although I volunteered to be an instructor, went through the two- day in-services and considered myself very AIDS aware, these things didn't really prepare me to teach the 27 new teenage faces that week. Why would they want to open up to a

> stranger? We didn't have time to meet with the teachers of the classes we presented to beforehand. I mean, consider this, in the previous Friday's homeroom, the kids probably talked about the vacation schedule. On Monday, we began teaching what one students described as, 'that AIDS stuff.' So many of us went in cold. In time, as we got to know the students, it became easier to talk with them and get them involved in the activities we had planned. But we still only had 20 minutes each morning. (District C AIDS teacher 7, interview, January 21, 1994).

The uneasiness with the process of instruction emanated primarily from the framework in which the AIDS instruction was provided in District C. With the curriculum being delivered in the homeroom setting, and not necessarily by homeroom teachers, the AIDS instructors found themselves unaware of the particular nuances of the classes. While some presenters taught classes that were more inhibited, when encountering their own classes, a camaraderie had often been established and dialogue developed freely. One teacher said:

> One topic I got asked about was condoms. Now that's something we really aren't supposed to talk about. The kids want to know what kinds there are. They want to know where to get them. They want to know if they can fully prevent AIDS. These are touchy subjects because I know they are not part of our district's curriculum. I worry about how my responses could easily be misconstrued and shared with a parent who may not want to ever hear the word 'condom.' On one hand you feel like you are withholding vital information, information that could save a life. So I tell them that they can be bought at any supermarket. But I know I must drive the discussion back to abstinence. I felt guilty about my response. (District C AIDS teacher 6, interview, January 20, 1994).

Other AIDS teachers experienced similar limitations. Many felt they had to develop quick responses to questions they were not prepared for. Curriculum restrictions also left exposed content areas which had not been previously discussed or resolved in teacher-training activities. The issues surrounding sexual expression, both heterosexual and homosexual, were prime examples of topics of interest to secondary students, even though the AIDS teachers did not have free reign to discuss them. While one AIDS instructor would choose to unabashedly discuss such topics, another would circumvent the issues:

> During one of my classes, the students [sophomores] wanted to know why homosexuals aren't just quarantined. 'Fags started it!' they say. This kind of thinking is tough to accept without a discussion, at least for me. I reminded them it isn't just homosexuals who transmit the disease. There are different theories why the disease struck this group first. Then we talked about other carriers, heterosexuals, the carriers who are asymptomatic and may be infecting others and not even know about it. We discussed the issues surrounding the latency period. I asked them what would be the point of isolating people? Who would pay for it? I reminded them of the situation in Africa, where the disease has affected heterosexuals. I tried to impress upon them that we need to identify and stop risk behaviors, instead of blaming groups. Maybe I was defending gays, their lifestyle, which is something we're not supposed to do, according to the state and district guidelines. But I felt the students' misconceptions were so great that I couldn't let them go unchallenged (District C AIDS teacher 5, interview, October 20, 1993).

> In my class of juniors, they were very vocal. One guy in the back said something to the effect of 'Kill all fags, let 'em die anyway, it's their fault.' Then someone else would say, 'It's not just fags who get it. Look at Magic Johnson. He got it.' And at that point it was time for me to stop the discussion, at least at that level. I told the class we don't know where the disease originated. Then I moved on to another topic. Afterwards I knew what the group really was interested in, and perhaps could have benefited from, was a discussion about the different types of sexual expression, that it's not just a gay versus straight morality issue. But I wasn't allowed to talk about

> that because maybe it would seem I was promoting the gay lifestyle. So I didn't pursue the topic any further (District C AIDS teacher 3, interview, December 13, 1993).

It soon became apparent to the District C AIDS teachers that a discussion of monogamy could not be sustained without a discussion of relationships and how sexual intercourse becomes a part of them. The teachers discovered that students had not given these issues much consideration. Student views often emanated from some personal experience, television or films and experiences they had witnessed in their own families. And it were these views which caused discussions to stray away from the abstinence message. One teacher recalled:

> To be honest, when I began the abstinence discussion, one female student mentioned her 16-year-old sister just had a baby. Another student added that her 17-year-old cousin had a baby, too. It was then that I realized the topic of abstinence certainly wasn't appropriate for a number of my students. After all, their peers had children. What's stopping them from becoming involved? But I had no idea how to address them, with the abstinence message looming over my head and their comments suggesting they needed to discuss something else (District C AIDS teacher 4, interview, December 14, 1993).

> My students were pretty up front when it came to the discussion on monogamy. I mentioned to them that even though they may be with one boy or girlfriend doesn't necessarily mean that you were their first. We call this 'serial monogamy.' We talked about what they thought makes a good relationship. They said things like commitment, being able to share feelings, having things in common. Abstinence is a hard sell (District C AIDS teacher 1, interview, December 3, 1993).

By the week's end, the District C AIDS teachers acknowledged that while the delivery of the curriculum did pose certain challenges, they felt confident their efforts were, for the most part, important. They were, however, uncertain about their effectiveness.

## FINDINGS

Principal findings reveal the impact of the federal and state governments' role in the development and implementation of AIDS educational policies. Federal policies designed to give direction to the country's AIDS education efforts have been slowed because of the conflicting views of morality held by policy makers who risk offending constituents. Many constituents fear that frank AIDS and sex education curricula will encourage promiscuity and illegal behaviors. This notion, while unsubstantiated, has persisted throughout the history of sex education in the United States (Brandt, 1987).

The federal government, through the Centers for Disease Control, require that the content of the nation's AIDS education efforts be determined locally and reflect community values. In order to receive federal funding, which is the only source of funding for AIDS education for the majority of the states, community review boards are required to take part in local policy and curriculum development, but the boards need not include representatives from at-risk groups. This emphasis on local control and community values reflects the inclusion of restrictive sex education principles which were established by President Ronald Reagan's Domestic Policy Council in 1987.

Funding for state AIDS education programs has been further complicated by amendments attached to federal appropriation measures at the insistence of conservative legislators which require that no federal funds be used to "promote homosexuality." The CDC, reacting to this legislation, has adopted regulations that prohibit federal funds from being spent on AIDS education materials which may offend some members of the community, even if the materials are not targeted to those parties who might be offended.

With the use of illicit drugs and non-prescription syringes unlawful in all states and sodomy illegal in 25, it is unlikely that the federal government will challenge the laws established by the states and endorse an educational policy which contains material which may contradict these laws.

Despite the emphasis of AIDS educational policy on abstinence by federal,state and local curriculum review boards during the first decade of the epidemic, the CDC-itself has reported that the number of AIDS cases has increased most rapidly among adolescents, young adults and women through heterosexual transmission (1993). However, the absence of an effective national AIDS education policy has not been recognized as contributing to the country's inability to contain the spread of the disease.

Arizona adheres to the CDC's AIDS education policies and encourages school districts to develop educational materials which reflect the values and culture of their local communities. The state's AIDS Curriculum Review Board developed an education program which reflects, for the most part, the values of the dominant, conservative community. What curricula does not reflect the diversity of class, race, ethnicity, gender or sexual orientation of the broader community. In fact, teachers related that the curriculum they delivered to students did not adequately address explicit issues and needs often raised by the diverse students they taught in the classroom.

Principals have been slow to respond to the mandate, and did not consistently encourage teachers to attend AIDS education in-service training sessions, deliver instruction, and were lax in the arrangement of follow-up AIDS education activities. With no compliance mechanism in place, some principals did not perceive the need to act beyond what was minimally required of them by the state and the district. If their school's AIDS education effort consisted of showing a single video, they felt they were in compliance.

However, evidence was found that a handful of principals chose to make AIDS education a priority at their local school sites. They perceived the severity of the epidemic and supported AIDS education efforts in their school prior to the passage of the mandate, granted release time for teachers to attend conferences and even attended in- services themselves.

Nevertheless, principals and teachers alike recognized the additional demands placed upon them in an already crowded curriculum. Many teachers were reluctant to serve as their site's AIDS educator. Those who did volunteer acknowledged a personal commitment to the issue and to their students. They provided instruction with limited resources available to them and with minimal training. Often times the AIDS instructors debated internally about which topics to discuss with students, topics deemed taboo by Arizona's AIDS education policy standards. Many teachers chose to discuss explicit issues with their students despite the restrictions of the policy.

For those teachers unhappy about having to provide AIDS instruction, the knowledge they did deliver to students could easily be controlled. Abbreviated forms of the curriculum were presented which emphasized that unless abstinence and heterosexuality are adhered to, death is certain and deserved. By providing such fragments of the curriculum, the demands on the teachers remained minimal and manageable, especially for those who were uncomfortable or unfamiliar with the AIDS/HIV curriculum.

As the teachers sought to find ways to reconcile the state mandated AIDS education policy with their own beliefs and value systems intact, it became clear that many found it difficult to reconcile what the state was asking of them, and what kinds of information the students requested and needed. For the most part, the teachers had to use their own personal and professional judgment when determining what kind of information to discuss with their students. Often times, teachers responded with a rationale or defense when describing the curriculum they delivered which exceeded the terms of the mandate.

In the end, this educational policy was transformed then, into a series of personal, unofficial guidelines and coping strategies created and controlled by the practitioners, in this case, the teachers. The federal and state mandated pieces of the puzzle were jammed into place by the policy makers who were eager to legislate with values in hand, without an understanding of the ramifications of the issue, adequate resources, compliance mechanisms, and without a complete awareness of what is being asked of the practitioners.

Setting the stage for a formal AIDS education program in the Arizona public schools was the 1991 passage of the AIDS education mandate. While previous efforts sought to enact an AIDS education mandate in a variety of different forms, final passage of the mandate became a reality after compromises were struck among stakeholders concerning such issues as the role of the schools, the political climate and the inclusion of anti-homosexual language. These compromises were the result of the actors taking into account what other stakeholders were doing or were about to do, in this case, constituents.

However this was hardly possible in the case of the development of the Arizona AIDS K-12 mandate, where those most versed in confronting the epidemic, gay AIDS social service agencies and organizations representing risk groups, such as Latinos and African-Americans, were excluded.

Prior to the passage of the mandate, Districts A and B did not have a formal AIDS education program in place. Once the mandate was approved, even with its carefully constructed language, getting district A and B teachers to attend and participate in the AIDS training in-services proved to be difficult. Principals did not consistently encourage teachers to attend training in-services, or were lax in the arrangement of follow-up AIDS education activities for the next academic year. With the support of their principals, district C teachers were quick to respond to the AIDS crisis by organizing its own district-wide curriculum committee and instructional strategies before the state of Arizona had even passed the mandate.

District A and B K-5 teachers who attended AIDS in-services emphasized hygiene and basic health skills, as required by the mandate. Some teachers at this level, however, simply chose not to deliver the AIDS education curriculum because they either chose not to attend an in-service, felt it was another curriculum task (on an already full-plate) that they were not being compensated for, or argued that any discussion of sexuality was against their personal values. As a result, mandated instruction was transformed then, into a series of personal, unofficial guidelines and coping strategies created and controlled by the practitioners, in this case, the teachers.

District B and C middle school and secondary teachers sought to find ways to balance the state mandated AIDS education guidelines with their own beliefs and value systems, it became clear that many found it difficult to reconcile what the state was asking of them, with the kinds of information the students requested and needed. for the most part, the teachers had to use their own personal and professional judgment when determining what kind of information to discuss with their students. Often times, teachers responded with a rationale or defense when describing curriculum delivery which was either not acceptable under the terms of the mandate, or was simply inaccurate or incomplete.

All of the teachers appeared fearful of challenging the tenets of a mandated AIDS educational policy, and having their instruction be misconstrued by students who might relay that information to conservative administrators and parents. Other teachers felt constrained by the mandate's guidelines because they felt it was out of touch with student needs. Other teachers acknowledged that they delivered the curriculum poorly because they felt inadequately prepared, were uncomfortable about teaching material they received in "a crash course," felt the content did not reflect their personal value system, or believed schools should not be held accountable for provided that "should be taught at home."

Also evident were those teachers who were confident in their ability to engage students in discussions which, at times, strayed from the topics deemed acceptable by the mandate. Working around and within the limitations of a restrictive policy with keen communicative skills became a pedagogical technique several of the secondary AIDS educators appeared to be quietly proud of.

Several participants concluded that having an educational policy in place "was better than not having one," since the AIDS curricula probably would not be delivered to students at all. Also evident were proactive administrators and teachers in District C were able to anticipate students' needs. Evidence of "pre-mandated" AIDS education programs could be found in District C schools in which the participants felt a personal belief in the importance of the issues at hand. Networks of educators interested in clarifying, developing, an implementing a policy collaboratively allowed one school district to respond more quickly to the mandate than others.

In the end, it can be said that without an understanding of participant perspectives, of the educators who work directly with students, that the efforts of policymakers will never rise above the symbolic. Mandated, restrictive policies imposed from above without dialogue, adequate training, resources or accountability, only succeed in alienating the practitioners and ultimately, failing to meet the needs of all student.

## REFERENCES

Aiken, J. (1987). Education as prevention. In H. Dalton (Ed.), *AIDS and the law* (pp. 90-95). New Haven: Yale University Press.

Arizona Department of Education. (1992, September). *HIV/AIDS education program*, 2.

Brandt, A. (1987). *No magic bullet: a social history of venereal disease in the united states since 1880*. Oxford: Oxford University Press.

Centers for Disease Control. (1988). Guidelines for effective school health education to prevent the spread of AIDS. *Morbidity and mortality weekly report*, 37.

DiClementi, R. (1992). *Adolescents and AIDS: A generation in jeopardy*. Newbury Park: SAGE.

Dodds, S., Volker, M. & Viviand, H. (1989). Educating children and youth about AIDS. In J. Seibert & R. Olsen (Eds.) *Children, adolescents, and AIDS*, (pp.119-146). Lincoln: University of Nebraska Press.

Eckland, J. (1989). Policy choices for AIDS education in the public schools. *Educational evaluation and policy analysis*, 11, (4), 377-387.

General Accounting Office. (1990). *AIDS education: public school programs require more student information and teacher training*. (GAO/HRD 90-130). Washington, DC: Author.

Nadel, M. (1990). *AIDS Education: Gaps in Coverage Still Exist*.(GAO/T-HRD-90-26, Statement of Associate Director for National and Public Health Issues, Human Resources Division before the Senate Committee on Govermental Affairs).

National Association of State Boards of Education. (1990). *AIDS, HIV, and school health education: state policies and programs 1990*, pp. 12-13. Alexandria, VA.

Popham, J. (1993b). Wanted--AIDS education that works. *Phi delta kappan*, 559-562.

Sex information and education council of the United States. (1991). *Winning the battle:developing support for sexuality and HIV/AIDS education.*

World Health Organization. (1993, January). *Weekly epidemiological record, 68*, (3), 9-10.

## Appendix A--Interview Protocol

1. What is the Arizona AIDS education policy?
2. How was the policy developed into its present form?
3. Who is responsible for implementation?
4. What can be taught about AIDS? How was this determined?
5. Who decided at the federal level?
6. Who decided at the state level?
7. Who decided at the district level?
8. What learning outcomes are expected at the state level?
9. How is the implementation of the mandate being funded?
10. How are districts delivering the curriculum?
11. How are schools delivering the curriculum?
12. What kinds of instructional obstacles have arisen?
13. How have educators confronted them?

14. How could the curriculum and pedagogy be improved?
15. How is compliance being monitored?
16. How is instruction being evaluated?
17. What occurs if a parent, student or staff member does not want to participate in the AIDS education unit?

## About the Author

Grace C. Huerta, Ph.D.-- Assistant Professor
Department of Secondary Education,
College of Education, Utah State University
Logan, UT 84322-2815
graceh@cc.usu.edu *(801) 797-3946*

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@pisces.rutgers.edu*

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Sherman Dorn
*sfxj9x@scfn.thpl.lib.fl.us*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

**Volume 4 Number 14** **August 27, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# Actual Schools, Possible Practices: New Directions In Professional Development

**Rebecca Novick**
**Northwest Regional Educational Laboratory**

*novickr@nwrel.org*

**Abstract** There is increasing recognition that school reform and staff development are integrally related. Yet, despite a rich literature on adult learning and human development which supports teachers' need for a wide array of opportunities to construct their own understandings and theories in a collaborative setting, top down mandates have frequently left teachers out of the reform process. It is argued here that effective staff development should be tied directly to the daily life of classroom and grounded in the questions and concerns of teachers. Both a theory of pedagogy that advocates teaching for understanding and learning as understanding and a model of staff development based on practical knowledge enriched by critical reflection are discussed.

Education is about learning how to deal with uncertainty and ambiguity. It is about learning how to savor the journey. It is about inquiry and deliberation. It is about becoming critically minded and intellectually curious, and it is about learning how to frame and pursue your own educational aims. It is not about regaining our competitive edge (Eisner, 1992).

Ever since the authors of *A Nation at Risk* (1983) warned that a rising tide of mediocrity in our educational system was compromising our nation's ability to be competitive in the world economy, education reform or restructuring has been proposed, not only to improve schooling, but as the solution to our nation's ills. Yet there is considerable agreement that these sometimes conflicting waves of reform have produced disappointing results (Clark & Astuto, 1994; Darling-Hammond & McLaughlin, 1995). And although it is commonsensical that good schools need excellent teachers, teachers have often been excluded from the process, both of planning reforms and the professional development opportunities necessary to implement them (Lieberman, 1995). As early as 1957, the National Society for the Study of Education recommended that schools and entire staffs become collaborators in providing inservice education. However, Sykes (1995 ) points out that over 40 years later, "teachers are frequently the targets of reform, but they exert relatively little control over professional development" (p. 465).

In the *1988 Annual Report of the Carnegie Endowment for the Advancement of Teaching* (Boyer, 1988), Boyer reported that morale within the teaching profession had substantially declined since the publication of *A Nation at Risk*, that in fact, teachers were "demoralized and largely unimpressed" by the reform actions taken in the previous five years. Since that time, the tension between old and new waves of reform (Hargreaves, 1994) and the "policy collisions" between them (Darling-Hammond, 1990) have, in Darling-Hammond's words, sometimes "created an Alice in Wonderland world in which people ultimately begin to nod sometimes "created an Alice in Wonderland world in which people ultimately begin to nod blithely at the inevitability of incompatible events" (p. 344). In such a climate of confusion and contradiction, and with little input into the reform process, it is not surprising that many teachers have opted to close the classroom door and wait for it all to go away.

Recently, however, there has been increasing recognition that teachers and teachers' knowledge gained from and embedded in their everyday work with children should be at the center of reform efforts and professional development activities (Darling-Hammond, 1994; Lieberman, 1995). It is that model of professional development which is advocated in this paper. At the heart of the dialogue regarding school reform and professional development are questions regarding the nature of learning and the purposes of schooling. In the next section, these questions are explored.

## Learning In Our Nation's Schools: Simple-Minded or Muddle-Headed?

Legend has it that during a heated philosophical argument, Bertrand Russell announced to his protagonist and teacher, Alfred North Whitehead, "This issue cannot be resolved. The problem is that I am simple-minded and you are muddle-headed." In many ways, the dialogue over school reform and the role of teachers in such reform has reflected this dilemma.

Our educational system has drawn heavily on theories of behaviorism and the scientific management ideas of Frederick Taylor. The positivist assumptions of objectivity, rationality, efficiency, and accountability have exerted a strong influence on our curriculum, assessment, and classroom climate. Skills are regarded as the sum of their component parts, often taught directly and practiced in isolation from their use before being brought back to the whole (Crawford, 1995). In the "transmission" or behaviorist approach to education, the teacher's job is the direct instruction of information and rules.

Implicit in this view is the image of the learner as passive, a vessel to be filled with knowledge by the teacher. Because our educational system frequently reflects the assumption of hierarchical intelligence (Darling-Hammond, 1994) in which, as Meier (1995) notes, the top does the critical intellectual work and the bottom is left with doing the daily 'nuts and bolts' or 'how-to' (p. 369), teachers are often viewed as technicians, purveyors of a "canned curriculum" provided by a very powerful knowledge industry (Goodman, 1994). In the best tradition of scientific management, the classroom has been frequently portrayed as a factory and children regarded as products to be produced as efficiently and systematically as possible.

Interacting with and complementary to this approach is a psychometric philosophy of education, which posits that the learner possesses measurable abilities; individual differences in performance are regarded as reflecting differences in amount of ability (Elkind, 1991). In a psychometric approach, education is seen as imparting quantifiable knowledge and skills which can be measured objectively on standardized tests. Answers are either right or wrong, and subjects are autonomous, with each discipline possessing its own scope and sequence of skills. Learning is viewed from this very linear perspective, "much like a train racing along a railroad track" (Wills, 1995).

The course is predetermined and no detours are allowed. The only variable is the speed by which the journey is made. An unusually quick trip denotes a child whose learning ability is above grade level; an on-time arrival denotes a child at grade-level. All educators are familiar with the many labels for those who arrive late. Of course, many of those late arrivals never complete the trip, eventually choosing to jump from the train (p. 262).

### Development as the Aim of Education

Over the last half century, research from a variety of disciplines has provided support for other approaches to education that are responsive to how children learn and develop. Variously referred to as "teaching for understanding" (Cohen, McLaughlin & Talbert, 1993), culturalism (Bruner, 1996), developmentally appropriate practices (Bredekamp, 1987; Bowman, 1994), and the transactional model (Weaver, cited in Braunger, 1995), these approaches draw on the theories of Piaget, Dewey, and Bruner, and Vygotsky.

Representing the disciplines of education, cultural anthropology, and psychology, these theorists propose an integrated, holistic approach in which learning is viewed as an active theorists propose an integrated, holistic approach in which learning is viewed as an active process, driven by the innate need of children to make meaning of their experiences. Children, rather than receiving meaning from expert adults, construct and negotiate knowledge and understanding through interaction with the social and physical environment. Thus, learning is regarded as a process, the personal discovery of the learner of the meaning of events for him or her. Each new discovery changes or refines prior knowledge, building a complex network of interconnected concepts (Kostelnik, 1992).

Young children, in particular, need to establish a rich, solid conceptual base from which all future learning will proceed (Kostelnik, 1992). Such a base enables children to make sense of their experience by forming connections between what they know and understand and the knowledge and concepts encountered in the new environment. Without this base, learning facts and isolated skills may resemble nonsense-syllable learning, often quickly mastered and just as quickly forgotten. Early childhood educators are concerned that children have the capacity and opportunities to use their knowledge and skills within the context of meaningful activities, both inside and outside the classroom. As Doris Lessing has observed, true learning is understanding something on deeper and deeper levels.

Although followers of Piaget have emphasized the child's individual construction of knowledge, due to increasing attention to Vygotsky's theoretical framework, educators are beginning to understand that "making sense" is a profoundly social process, one in which culture and individual development are mutually embedded (Bowman & Stott, 1994). Because the child is viewed as intrinsically motivated, self-directed, and actively involved in the learning process, the role of the teacher, rather than dispenser of information, has been described as a planner of possibilities, a guide, ethnologist, researcher, and co-constructor of knowledge (Malaguzzi, 1994; Phillips, 1993).

In this view, although "teaching as telling" (Lieberman, 1995; Meier, 1995) is still a part of the educational process, it is only a part. As Bruner (1996) observes, "Even if we are the only species that 'teaches deliberately' and 'out of the context of use,' this does not mean that we should convert this evolutionary step into a fetish" (p. 22). Rather, learning is regarded as an adventure in which both teacher and children are engaged in joint inquiry, with teachers facilitating children's learning through "posing questions, challenging students' thinking, and leading them in examining ideas and relationships" (Cohen, McLaughlin & Talbert, 1993, p. 1). Children are encouraged to learn from and with each other in classrooms and schools that help children learn, in Eisner's words (1991), "to develop an ethic of caring and create a community that cares."

## Dangerous Dichotomies

While behaviorist approaches are characterized by teacher-controlled learning, instructional technology, quantifiable predetermined outcomes, and predictability, the transactional philosophy is characterized by following the child's lead, a "constant interchange of thoughts and ideas" (Kostelnik, 1992) and ambiguity. According to Elkind (1991), "The developmental approach tries to create students who want to know, whereas the psychometric approach seeks to produce students who know what we want" (p.9).

Polarized in this way, the dichotomies between traditional educational approaches and transactional approaches seem clear: product versus process, skill versus meaning, objectivity versus subjectivity, a passive versus an active learner, parts versus wholes, simplicity versus complexity, and accountability versus fuzzy-mindedness. In short, to return to Russell and Whitehead's argument, often the debate can be seen as offering a choice between being simple-minded and muddle-headed.

The reality, of course, is more complex. If education was originally instituted to meet the needs of the work place for a well-disciplined, homogeneous, semi-literate work force to "man" the factories and assembly lines, the employee of the twenty-first century, will be expected to be adept at finding, using, and making sense of information, problem-solving, thinking critically and imaginatively, resolving conflict, and understanding diversity. Clearly, in order to "produce" such a citizen and worker, skills and meaning, process and product, and parts and wholes are essential to the learning process. Students must be able to read, understand, and enjoy literature; be adept at solving math problems, and develop a positive attitude toward math, work collaboratively to solve problems and develop caring relationships.

attitude toward math, work collaboratively to solve problems and develop caring relationships.

Teaching, then, addresses all four components of learning identified by Katz (1988): knowledge, skills, dispositions, and feelings. The role of teachers, rather than as purveyors of a canned curriculum, is to start where the learner is, helping the learner to build new knowledge and understandings. When students are encouraged to ask meaningful questions and formulate alternative solutions, appreciate multiple viewpoints, and develop multiple intelligences, a certain amount of uncertainty and ambiguity are not only inevitable, but necessary for good teaching. A major goal of staff development activities must be to help teachers find their own balance between "coverage and making sense of things" (Meier, 1995), between getting children ready for next year" and encouraging what Malaguzzi (1994) refers to as "the hundred languages of children."

Yet, as Tyack and Tobin (1993) point out, our idea of a "real school" is remarkably resistant to change. The literature on school reform has focused on two issues in particular which challenge educators' ability to make education responsive to the needs of children and their families: evaluation practices and the marketplace metaphor of schooling (Eisner, 1992).

**Evaluation practices.** The belief that our faltering educational system is putting our nation at risk economically has gained popular appeal, resulting in the promotion of national and/or state standards and assessments as a means for improving curriculum and student performance in school. A number of educators and researchers, however, have raised serious concerns about "top-down specifications of content linked to tests" (Darling-Hammond, 1994, p. 478). For example, many educators argue that such attempts to "stamp a uniform education" (Bowman, 1994) on students leaves the learner out, making it hard for him or her to build new knowledge and new understandings (Goodman, 1994; Meier, 1995; Nieto, 1994). A 1992 study by Poplin and Weeres (cited in Nieto, 1994) concluded that students became more disengaged as the curriculum, texts, and assignments became more standardized. This is particularly true for poor and minority students, who often start out farthest from the standard and for whom "turning standards into simple yardsticks can be devastating" (Goodman, 1994, p. 39).

As long as our educational system considers coverage of a prescribed curriculum, mastery of discrete skills, and increase of achievement test scores of paramount importance, implementing a "mindful" (Bredekamp & Rosegrant, 1992) and "thinking" (Darling-Hammond, 1994) curriculum will remain problematic. Teachers striving to implement such a curriculum will often struggle to meet the requirements of two incompatible systems based on widely differing philosophies of education.

But how do we know that we are meeting valid educational goals? Whereas a number of educators are concerned that standards, based on in industrial model of schooling, with an emphasis on uniformity, can be harmful to teaching and learning, well-conceived curriculum standards can be used as "tools for informing curriculum building, teaching practice, and assessment" (Darling-Hammond, 1994, p. 488). According to Bredekamp & Rosegrant (1995), "well-developed national content standards would be advantageous for at least five reasons. They have the potential to provide the curriculum with important content, conceptual framework, coherence, consistency, and high expectations" (p. 9). Rather than creating a wall around the curriculum, such flexible standards can provide a framework for local educators to reflect on and evaluate their own efforts to change their teaching practices to better meet the needs of children and families in their own communities

**Nation at-risk or children at-risk?** Perhaps equally problematic for school reform efforts is the tension between the concept of education as a means to improve academic performance to make our country more competitive in a global economy and education as nurturing children's intelligence and ability to make sense of their experience. Tyack (1992) describes two current conceptions or versions of educational reform: a "nation-at-risk" model, or a "children-at-risk" model. In a nation-at-risk model, education is conceived, in Eisner's words, as "a competitive race, the front lines in our quest for international supremacy" (1991, p. 10). In a children at risk model, rather than increased competition between children and schools, the goal becomes meeting the health and social needs of an increasing number of children who are experiencing behavioral, emotional, and learning problems (Tyack, 1992).

Arguing that schools and communities are adversely affected by nonacademic problems among students and families, proponents of this view advocate for schools to establish links with community service providers as an essential component of restructuring schools to meet with community service providers as an essential component of restructuring schools to meet the needs of children and their families. In addition, schools are encouraged to create caring communities of learners and often, in Garmezy's words, "to serve as a protective shield to help children withstand the multiple vicissitudes that they can expect from a stressful world" (Garmezy, 1991). This view is in sharp contrast with the "back-to-basics" movement which seeks to reduce a school's purview to the instruction of children in the traditional "3-Rs," with a heavy emphasis on skill acquisition and memorization of facts.

If school reformers are to avoid the pitfalls both of Russell's and Whitehead's arguments and the Alice in Wonderland world described by Darling-Hammond (1990) in which conflicting mandates and expectations create confusion and stress for teachers and children, professional development activities will need to help teachers balance the inevitable tension between preparing children for the world of work and viewing education as lifelong learning and inquiry. To do so requires time for observation, reading, reflection, dialogue with colleagues, and support for these practices at the district, state, and federal levels. Wilson and colleagues (1996) note:

> If visions of reform hold any prospect of influencing American schools, new learning will need to occur at multiple levels. Policymakers will have to learn, as well as children; teachers, as well as parents. Administrators, curriculum developers, school board members - everyone will have to learn (p. 469).

### Professional Development and School Reform

Researchers on school restructuring have identified a number of commitments and competencies which lead to improved outcomes for children, including: (a) high expectations for all children (Newmann, 1993; Benard, 1993; Nieto, 1994); (b) a commitment to learn from and about children, building on the strengths and experiences which children bring to school (Bowman, 1994; Delpit, 1995; Ladson-Billings, 1995; Meier, 1995); (c) "giving wider choices and more power to those closest to the classrooms" (Meier, 1995, p. 373); (d) working collaboratively with families and the community; and (e) development of schools as caring communities (Lewis, Schaps, & Watson, 1995; Meier, 1995; Newmann, 1993), defined by Lewis, Schaps & Watson as: "places where teachers and students care about and support each other, actively participate in and contribute to activities and decisions, feel a sense of belonging and identification, and have a shared sense of purpose and common values."

But, as Joyce and Calhoun (1995) point out, "if a major dimension of schooling is creating caring communities for children, much less attention has been directed at how to develop schools as organizations that nurture the professionals who work within them" (p. 55). Despite a rich literature on adult learning and human development which supports teachers' need for a wide array of opportunities to observe, read, practice, reflect, and work collaboratively with peers, the "one-shot workshop" remains the primary method of providing inservice professional development. As Miller (1995) puts it," The old model of staff development survives in a world where everything else has changed" (p. 1).

Institutions providing training and certification for teachers do not usually prepare them to create schools where dialogue, reflection, and inquiry are valued and practiced. Rather, teacher-preparation institutions typically use a model in which experts impart technical skills and knowledge to teachers in a context that is divorced from the classroom. Courses are organized according to academic disciplines, with scant attention paid to examining the problems of actual practice (Cohen, McLaughlin, and Talbert, 1993; Little, 1993). Not only are practicums and student teaching seldom supervised by the same people who teach the courses, but there is little institutionalized support for making the connections between what it means to understand a subject and how it can be taught and learned (Cohen, McLaughlin & Talbert, p. 45). When teacher preparation is based on a transmission model of learning, a central dilemma for teachers becomes how to teach in ways one has seldom or never experienced (Darling-Hammond & McLaughlin, 1995; Little, 1993; Meier, 1995).

### Inquiry Based Professional Development

A new kind of structure and culture is required, compatible with the image of "teacher as intellectual" rather than teacher as technician. Also required is that

> "teacher as intellectual" rather than teacher as technician. Also required is that educators enjoy the latitude to invent local solutions rather than adopt practices thought to be universally effective (Little, 1993).

New approaches to professional development have emerged from the Weberian tradition that emphasizes "verstehen," the interpretative understanding of human experience and information (Bogdan & Biklen, 1982). The "interpretative turn," which began in the last half of the nineteenth century, first expressed itself in drama and literature, then in history, then in the social sciences and epistemology, and finally in education (Bruner, 1996). This influence is reflected in the increased appreciation for practical knowledge enriched by critical reflection. Bruner notes, "The object of interpretation is understanding, not explanation; its instrument is the analysis of text. Understanding is the outcome of organizing and contextualizing essentially contestable, incompletely verifiable propositions in a disciplined way" (p. 90).

**Teaching for understanding.** Proponents of a transactional approach are firmly committed to both teaching for understanding and learning as understanding. As early as 1967, Schaefer proposed that schools should be centers of inquiry "where faculties continuously examine and improve teaching and learning and where students study not only what they are learning in the curricular sense, but also their capacity as learners" (cited in Joyce & Calhoun, 1995, p. 51). If the preferred pedagogical mode of behaviorism is skill and drill, in the transactional approach, collaboration and dialogue provide a large part of children's and teachers' learning opportunities.

In such schools, teachers, often in concert with parents and children, engage in inquiry into curriculum, instruction, and assessment in efforts to improve teaching and children's outcomes. As teachers collaborate to develop and evaluate new practices, such as authentic assessment, a literacy program, or multiage classrooms, the inquiry process itself becomes an important component of staff development, providing opportunities for teachers to articulate goals, address questions and concerns, and find solutions together (Clark & Astuto, 1994; Darling-Hammond & McLaughlin, 1994).

Unlike standardized curricula, which provide certainty and predictability, new approaches to teaching require teachers to weigh conflicting demands and reflect on their own practices. Researchers have consistently found that in order for teachers to facilitate higher order thinking in children, they too must have ample opportunities to construct their own understandings and theories. As Joyce and Calhoun (1995) point out, "staff development must not be offered as, "Here is stuff that has been researched, so use it!" (p. 54). Rather, effective staff development requires opportunities to be enriched by what Meier (1995) refers to as "the power of each other's ideas." In a study of nine Northwest schools, (Novick, 1995) a consistent theme was the need for curriculum review and collaborative study at the building level. All sites found that, as the research shows, simply implementing what others have deemed as "best practices" does not lead to a sense of competence, purpose, or commitment, essential to the implementation of a "mindful" curriculum. As Fullan (1993) observed, "It's not a good idea to borrow someone else's vision." Thus, a certain amount of "reinventing the wheel" was considered a vital part of staff development by these educators.

Peer coaching and mentoring. Peer coaching provides additional avenues for teachers to share expertise perspectives, and strategies with each other. Cohen, Talbert & McLaughlin (1993) point out that "understanding teacher-thinking involves understanding how teachers respond to an ever-changing situation with knowledge that is contextual, interactive, and speculative" (p. 55). For this reason, they advocate that teacher development programs be structured around peer coaching or mentoring in which the relationship between learner and coach is grounded in actual classroom practice. Learning new practices often involves changing old habits that have made teaching comfortable and predictable. Because teachers have to both *learn* new habits and *unlearn* old ones, as one teacher put it, "The comfort is for *not changing*" (Cohen, McLaughlin & Talbert, 1993, p. 93). This teacher contrasts ongoing peer coaching with the typical inservice workshop experience:

> I think you need the support of people with new ideas. The only way we change our teaching is to talk to people who are also changing. And you need time to talk to one another. But not on just a one-time basis, for it's got to be reoccurring. If Suzanne (a teacher educator) had come into my room and done a couple of lessons and said,

> teacher educator) had come into my room and done a couple of lessons and said, 'Okay, this is the way you teach,' I would not have changed. But because this has been ongoing for several years, I really am seeing changes in myself - in the way I think. It is because of that support of talking with her and Carol Miller (a fellow teacher) (p. 93).

Such mentoring relationships in which both teacher and coach view themselves as learners can be set up both inside and outside the school. For example, since the late 1980s, more than 20 Professional Development Schools (PDS) have been created for the purpose of enabling veteran and novice teachers to work together. Many of these partnerships are connected to major reform networks such as the Coalition of Essential School and the Comer School Development Program, noted for their innovative and successful practices. In such partnerships, both novice and experienced teachers benefit from the relationship as they engage in discussion, joint inquiry, and action research (Darling-Hammond and McLaughlin, 1995).

The types of networks and partnerships in which schools engage are determined by the changing needs of teachers and children. Darling-Hammond and McLaughlin (1995) suggest: "What does need to be a permanent addition to the policy landscape is an infrastructure or "web" of professional development activities that provide multiple and ongoing occasions for critical reflection and involves teachers with challenging content" (p. 600).

**School/university partnerships.** University/school partnerships can provide ongoing opportunities for teachers to discuss research and practice and to engage in professional development which is grounded in teachers' experiences. In addition, these partnerships can provide opportunities for teacher-educators to teach in ways that encourage inquiry into educational practice. Goodlad (1994) notes, "It is unrealistic to expect teachers to create schools for inquiry when the settings in which they are prepared are rarely reflective" (p. 18). Reciprocal school/university relationships can help solve the riddle posed by Meier (1995): "We cannot pass on to a new generation that which we do not ourselves possess" (p. 146).

In Oregon, Portland State University, in partnership with three selected local school districts and Education Service Districts, has developed an off-campus masters program for practicing teachers designed as critical inquiry into educational practices and their relationship to school reform. Co-taught by a Portland State University staff member and an instructor from the district office, teachers are encouraged to reflect on their own personal experiences and issues and concerns regarding their own teaching in group discussions and in a learning log or journal.

Portfolios with scoring guides provide the major evaluative tool; and the masters thesis consists of an action research project conducted by teaching teams. In this way, as one district staff development coordinator who has served as instructor for one of the three programs put it, "You're not just piling up courses and when you get to the end, you're just relieved to get your degree." Instead, the educational program utilizes a constructivist approach in which "teachers reinvent curricular theory for themselves."

Over a two-year period, teachers participating in the program meet over 40 outcomes in four major content areas, including teaching and learning, inquiry for school improvement/change, social and cultural issues, and interpersonal skills to effect educational change. In order to create an integrated curriculum, all four content areas are woven through all courses. According to the district staff development coordinator quoted above, "Every quarter consists of collaboratively inventing a course of study that is unique. It has been exhausting, but is the most exciting staff development I have ever been involved in."

**Teacher networks.** In Montana, three school districts have formed a partnership in order to provide "ongoing professional development that is an integral characteristic of schools as communities of learners" (Mission Valley Consortium, 1995, 96). Based on the premise that "conversation, reflection, and continuous improvement" are essential for effective staff development, the consortium offers staff development opportunities that "provide a common direction, yet allow individual building staffs to design professional development plans unique to their own needs and interests" (Mission Valley Consortium). Parents are invited to participate in individual schools and with the Consortium at large.

Study groups, workshops, and courses for credit sponsored by the Consortium have included the following areas of study: Assessment; Children and Society; Cognition;

included the following areas of study: Assessment; Children and Society; Cognition; Cooperative Learning; Developmentally Appropriate Curriculum; Inclusion; Integration of Curriculum; Renewal and Leadership; Teaching and Learning; and Technology. Not only have standardized test scores improved, but, as the Consortium Catalogue notes, the consortium acts as a "positive persistent disturbance" in the process of change:

> Despite the many challenges of improving schools, we are seeing our faculties move toward a more constructivist approach to teaching and authentic forms of assessing learning. Without a doubt, all of us have increased our conversation about curriculum, learning, and children, and we believe that it is through this increased conversation and collaboration that significant and sustaining change will occur.

Lieberman (1995) cites two examples of teacher networks: The Foxfire Teacher Outreach Network and the Four Seasons Network. The Foxfire Network is an example of a network created by teachers for teachers, having grown out of one teacher's struggle to interest his students in learning in his English class. Initially, teachers were invited to participate in classes over the summer where they learned strategies such as encouraging students to choose their own topics and identify their own learning needs with teachers serving as guides. Currently, more than 20 groups of teachers meet throughout the school year to reflect on practice.

The Four Seasons Network brings together teachers from three reform networks: The Coalition of Essential School, the Foxfire Network, and Harvard University's Project Zero. Organized by the National Center for Restructuring School and Teaching (NCREST), the purpose is to support and encourage teacher participation and leaderships in the area of assessment (Lieberman, 1995). After initially participating in two summer workshops, year-round support is provided through the use of an electronic network. Through on-going access to new ideas in a supportive community, teachers are able to serve as catalysts for change in their school and classrooms.

**Collaboration with early care and education providers.** Collaboration with early care and education providers is an important aspect of providing continuity for children as they make the transition from preschool to kindergarten. In addition, engaging in collaborative professional development activities can be mutually beneficial to elementary school teachers and preschool and childcare providers: early care providers bring a rich experience with active, engaged learning, collaboration with families, and cultural pluralism (Phillips, 1994); elementary teachers draw on a more formal education in curriculum, instruction, and assessment.

Yet, due in part to our strongly held beliefs that the early care and socialization of children is not only the right, but is also the responsibility, of the family, our child care and preschool systems have never been integrated into a comprehensive educational system (Kagen, 1991). Isolated from the educational mainstream, as well as from each other, there is typically little networking between preschool and kindergarten programs (Love, Logue, Trudeau, & Thayer, 1992). Differences in status (teaching versus babysitting) and remuneration (child care providers often receive poverty-level wages) may militate against open communication.

During the last 10 years, however, the National Association for the Education of Young Children (NAEYC) has engaged in a number of activities to foster professional identity and visibility for the field of early childhood, including publishing guidelines for developmentally appropriate practice (Bredekamp, 1987), and more recently, a conceptual framework for the professional development of early childhood educators (NAEYC, 1994). Kagan (1994) noted:

> Professionals in the field of early care and education have begun to take stock of their own situation: fragmentation of services; competition with colleagues for scarce resources, including space, staff, and children; discontinuity and isolation from mainstream services, often including schools; less than optimally effective training and advocacy; and inequitable and unjust compensation and benefits (Kagan, p. 186).

Increased communication between these two distinct realms and opportunities to engage in joint staff development activities can do much to help children and their families build on in joint staff development activities can do much to help children and their families build on the positive aspects of their experiences as they make transitions (Regional Educational Laboratories, 1996). In addition, teachers/caregivers for early care and education can apply lessons learned from the struggle of elementary educators for professional status and adequate remuneration to their own efforts to achieve recognition and equity (Phillips, 1994).

### Schools as Caring Communities

Collaborative inquiry can only thrive in a climate of mutual respect, interdependence, and trust. The factory-model school, with an emphasis on competition, hierarchical authority, and a view of teachers and principals as interchangeable parts, still exerts a strong influence on our educational system. However, based on a synthesis of literature about human growth and development, Argyris (cited in Clark & Astuto, 1994) concluded that hierarchical, bureaucratic work environments are more likely to lead to immature behaviors, such as passivity, dependence, and lack of self-control and awareness.

In contrast, schools organized as caring communities have been shown to foster a shared sense of responsibility, self-direction, experimentation, respect for individual differences, and high expectations (Clark & Astuto, 1994; Lewis, Schaps & Watson, 1995; Newmann, 1994). When school staff (including principals, certified staff, counselors, and family advocates), parents, and children build on their own experiences and knowledge in an atmosphere that is psychologically safe (Espinosa, 1992), everyone's learning is enhanced. Deborah Meier, former teacher/director of the highly effective and innovative Central Park East Schools, notes that "although trust takes a long time to build, it is the most efficient form of staff development" (p. 130).

Key to the establishment of a community of learners is a principal who encourages teachers to examine teaching and learning and implement ideas and programs that result from reflective practice (Reitzug & Burrello, 1995). Just as the role of the teacher is changing from dispenser of knowledge to children to "co-constructor" of knowledge with children, the role of the principal is evolving from direct instructional leaderships to the role of facilitator of group inquiry, "collaborative leader," liaison to the outside world, and orchestrator of decisionmaking (Wohlstetter & Briggs, 1994). A Northwest principal observed, "I no longer believe in school restructuring. I believe in changing adults. And adults change when they feel secure and can personally make decisions to do so" (Jewett & Katzev, 1993).

Issues of social justice and equity are at the center of this vision of school reform and professional development. Opportunities to engage in reflective analysis of practice should include encouragement of staff to examine their attitudes toward different ethnic, racial, gender, and social class groups (Banks & Banks, 1995; Delpit, 1995). Creating a democratic school community in which everyone is regarded as both a teacher and a learner helps all concerned develop the habits of mind and heart necessary to build a more just and caring society. Meier (1995) argues, "Public schools can train us for such political conversations across divisions of race, class, religion, and ideology. It is often in the clash of irreconcilable ideas that we can learn how to test or revise ideas or invent new ones." (p. 7).

### Barriers to Effective Professional Development

**Time and funding.** The process of changing one's practice is difficult and slow (Cohen, McLaughlin & Talbert, 1993; Espinosa, 1992), even when there is adequate time for ongoing peer coaching, self-reflection, and colleagial inquiry. Yet, time -- arguably one of the most critical elements of staff development -- is usually in short supply for teachers whose typical day, in Eisner's words, "isolates them from their colleagues and gives them scarcely enough discretionary time to meet the needs of nature" (p. 723). Cohen, McLaughlin & Talbert (1993) documented the partnership between two teachers and a college professor who taught part time in their classrooms:

> For years, Miller and Yerkes (the teachers) had had no time to breathe during their typical workday. Half serious, half joking, Yerkes told Wilson (the college professor) that the biggest delight of having her teach every afternoon was that there was time to go to the bathroom, to get a glass of water, to make a phone call. These little luxuries had been unknown to her, and were no small reward for the decision to

> little luxuries had been unknown to her, and were no small reward for the decision to collaborate (p. 92)

Because teaching is defined as "time on task" in a classroom setting, teachers in the U.S., compared to most European countries, have very little "released time" for staff development (Darling-Hammond, 1993). Darling-Hammond cites a 1986 study which found that schools spent less than one percent on professional development, a figure that is declining even further in the current climate of budget cuts for education and social programs. For example, in some Oregon school districts, cuts to professional development budgets of 50 percent are planned for the coming year. Meier (1995) compares the four weeks of staff development time that a Saturn plant in Tennessee provides for its workers to the one or two days a year of professional development that most teachers enjoy. Given the inevitable complexities encountered in the reform process and the inadequate time for staff development, it is no wonder that school reform has been variously compared to "driving while changing the tires (Meier, 1995), "the swamp" (Schön, 1987), "grinding down a glacier's mountainside of living ice" (Santa, 1995), and a "tidal wave" (Sykes, 1995).

**Isolation.** The egg-crate elementary school, where children are moved in batches through prescribed curriculum, still provides the framework for our educational system (Tyack & Tobin, 1993). In what has been popularly described as "the second most private act," teachers teach approximately 30 children in classrooms that are typically isolated from each other. As Darling-Hammond points out, "Almost everything about school is oriented toward going it alone professionally." Inside school, teachers are inclined to think in terms of "my classroom," my subject," or "my kids" (p. 601). Most teachers have little experience with helping peers grow professionally and find the role of "teacher of teachers" uncomfortable at first (Hoerr 1996).

Sharing problems and their solutions, collegiality, and collaborative inquiry are incongruent with bureaucratic principles of efficiency, authority, and procedural specificity, which still exert a strong influence on our public schools (Clark & Astuto, 1994). Thus, in addition to time to breathe and funding for a diverse menu of professional development activities, structures which promote changes in attitude and practice must be in place. These include a democratic governing body, a supportive administration, open door policies, team teaching, and opportunities for both small and large group collaboration with colleagues inside and outside the school.

### Summary

Although schools have traditionally been places where teachers engage in direct instruction of 30 children who work quietly at their seats, this model of "teaching as telling" is giving way to an approach based on a view of children as actively engaged in constructing their own understandings through interactions with the social and physical environment. If schools are to become exciting places for children to grow and learn, teachers, like children, need opportunities to become actively involved in their own learning process. Effective professional development, then, is grounded in the questions and concerns of those who work closely with children, and in Little's words (1993), "are intricately interwoven with the daily life of the classroom," p. 137).

In this approach to professional development, teachers are viewed, not as technicians, but as intellectuals (Giroux, 1988), teacher leaders, peer coaches, and teacher researchers (Lieberman, 1995). Ample opportunities for teachers to engage in reflective study of teaching practices, experimentation, collaborative problem-solving, and peer coaching in a supportive community of learners are essential.

## References

Banks, C. A. & Banks, J. A. (1995). Equity Pedagogy: An essential component of multicultural education. *Theory into Practice,* 34(3), 152-158.

Benard, B. (1993). Fostering resiliency in kids. *Educational Leadership,* November, 44-48.

Bogdan, R. C. & Biklin, S. K. (1982). *Qualitative research for education.* Boston, MA: Allyn & Bacon.

Bowman, B. T. (1994). The challenge of diversity. *Phi Delta Kappan*, November, 218-224.

Bowman, B. T., & Stott, F. M. (1994). Understanding development in a cultural context: The challenge for teachers. In B. Mallory & R. New (Eds.), *Diversity and developmentally appropriate practices.* New York: Teachers College Press.

Boyer, E. L. (1988). Annual report, Carnegie Endowment for the Advancement of Teaching. Princton, NJ.

Braunger, J. (1995). *Tensions to resolve: Improving literacy programs in the context of school reform.* Portland, OR: Northwest Regional Educational Laboratory.

Bredekamp, S. (Ed.) (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age eight.* Washington, DC: National Association for the Education of Young Children.

Bredekamp, S. & Rosegrant, T. (1995). Reaching potentials through national standards: panacea or pipe dream? In S. Bredekamp & T. Rosegrant (Eds.), *Reaching potentials: Transforming early childhood curriculum and assessment.* Washington, DC: National Association for the Education of Young Children.

Bruner, J. (1996). *The culture of education.* Cambridge, MA: Harvard University Press.

Clark, D. L. & Astuto, T. A. (1994). Redirecting reform: challenges to popular assumptions about teachers and students. *Phi Delta Kappan*, March, 513-520.

Cohen, D. K. , McLaughlin, M. L. W., & Talbert, J. E. (1993). *Teaching for understanding: Challenges for policy and practice.* San Francisco, CA: Jossey-Bass Publishers.

Crawford, P. A. (1995). Early Literacy: Emerging Perspectives. *Journal of Research in Childhood Education,* 10(1), 71-84.

Darling-Hammond, L. (1990). Instructional policy into practice: "The power of the bottom over the top." *Educational Evaluation and Policy Analysis,* 12, 339-347.

Darling-Hammond, L. (1993). Reframing the school reform agenda; Developing capacity for schools transformation. *Delta Kappan Phi,* 74(10), 752-761.

Darling-Hammond, L. (1994). National standards and assessments: Will they improve education? *American Journal of Education,* 102, 479-511.

Darling-Hammond, L. & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, April,* 597-604.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom.* New York: The New Press.

Edwards, C., Gandini, S., & Forman, G. (Eds.) (1993). *The hundred languages of children.* Norwood, NJ: Ablex.

Eisner, E. W. (1991). What really counts in schools. *Educational Leadership,* 48(5), 10-17.

Eisner, E. W. (1992). The federal reform of schools: looking for the silver bullet. *Phi Delta*

*Kappan*, May, 722-723.

Elkind, D. (1991) Developmentally appropriate practice: A case study of educational inertia. In S. L. Kagan (Ed.), *The care and education of America's young children: obstacles and opportunities.* Chicago, Ill: The University of Chicago Press.

Espinosa, L. (1992). The process of change: The Redwood City story. In S. Bredekamp & T. Rosegrant (Eds.), *Reaching potentials: Appropriate curriculum and assessment for young children.* Washington, DC: National Association for the Education of Young Children.

Fullan, M. (1993). *Change forces.* New York: Falmer Press.

Garmezy, N. ( 1991). Resiliency and vulnerability to adverse developmental outcomes associated with poverty. *American Behavioral Scientist,* 34(4), 416-430.

Giroux, H. A. (1988). *Teachers as intellectuals: Toward a critical pedagogy of learning.* Granby, MA: Bergin & Garvey.

Goodlad, J. L. (1994). *Educational renewal: Better teachers, better schools.* San Francisco, CA: Jossey-Bass Publishers.

Goodman, K. S. (1994). Standards, not! *Education Week,* September 7. 39-42.

Hargreaves, A. (1994). *Changing teachers, changing times.* New York: Teachers College Press.

Hoerr, T. R. (1996). Collegiality: A new way to define instructional leadership *Phi Delta Kappan*, January, 380-381).

Jewett,J., & Katzev, A. (1993). *School-based early childhood centers: Secrets of success from early innovators.* Portland, OR: Child, Family, and Community Program, Northwest Regional Educational Laboratory.

Joyce, B. & Calhoun, E. (1995). School renewal: An inquiry, not a formula. *Educational Researcher*, April, 51-55.

Kagan, S. L. (1991). *United we stand: Collaboration for child care and early education services.* New York: Teachers College Press.

Kagan, S. L. (1994). Early care and education: Beyond the fishbowl. *Phi Delta Kappan*, November, 184-187.

Kohlberg, L., & Mayer, R. (1972). Development as the aim of education. *Harvard Educational Review*, 42(4), 449-496.

Kostelnik, M. (1992). Myths associated with developmentally appropriate programs. *Young Children*, 47(40) 17-23.

Ladson-Billings, G. (1995). *The dreamkeepers: successful teachers of African American children.* San Francisco: Jossey-Bass Publishers.

Lewis, C., Schaps, E.& Watson, M. (1995). Beyond the pendulum: Creating challenging and caring schools. *Phi Delta Kappan*, March, 547-545.

Lieberman, A. (1995). Practices that support teacher development: Transforming conceptions of professional learning. *Phi Delta Kappan*, April, 591-596.

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129-151.

Love, J. , Logue, M. , Trudeau, J. V. , & Thayer, K. (1992). *Transitions to kindergarten in American schools.* Washington D. C., U. S. Department of Education.

Malaguzzi, L. (1993). For an education based on relationships. *Young Children*, 47 (1), 9-12.

Meier, D. (1995). *The power of their ideas: Lessons from a small school in Harlem.* Boston, MA: Beacon Press.

Miller, E. (1955). The old model of staff development survives in a world where everything else has changed. *The Harvard Educational Letter*, XI (I).

Mission Valley Consortium. (1995-1996). *Professional development catalogue.* Polson, MT: Mission Valley Consortium

NAEYC (1994). NAEYC position statement: A conceptual framework for early childhood professional development. *Young Children*, March, 68-77

*A nation at risk: The imperative for educational reform* (1983). Washington, D. C.: U. S. Government Printing Office.

Newmann, F. M. (1993). Beyond common sense in educational restructuring: The issues of content and linkage. *Educational Researcher*, 22(20, 4-13, 22.

Nieto, S. (1994). Lessons from students on creating a chance to dream. *Harvard Educational Review*, 64(4), 392-426.

Novick, R. (1996). *School-Based early Childhood Centers: Challenges and Possibilities.* Portland, OR: Northwest Regional Educational Laboratory.

Phillips, C. B. (1993). The hundred languages of children. *Young Children*, 49(1), 17-18.

Phillips, C. B. (1994). The challenge of training and credentialing early childhood educators. *Phi Delta Kappan*, November, 214-217.

Reitzug, U. C. & Burrello, L. C. (1995). How principals can build self-renewing schools. *Educational Leadership*, April, 48-50.

Regional Educational Laboratories' Early Childhood Collaboration Network, (1996). *Continuity in early childhood: A framework for home, school, and community linkages.* "work in progress."

Santa, C. M. (1995) . *Improving the literacy program: A journey toward integrated curriculum.* Portland, OR: Northwest Regional Educational Laboratory.

Schön, D. A. (1987). *Educating the reflective practitioner.* San Francisco: Josey-Bass.

Sykes, G. (1996). Reform of and as professional development. *Phi Delta Kappan*, March, 465-467.

Tyack, D. (1992). Health and human services in public schools: Historical perspectives. *The Future of Children*, 2 (1), 19-31.

Tyack, D. & Tobin, W. (1993). The "grammar" of schooling: Why has it been so hard to change? *American educational research journal*, 31(3), 453-479.

Wills, C. (1995). Voice of inquiry: possibilities and perspectives. *Childhood Education, Annual Theme*, 261-265.

Wilson, S. M., Peterson, P. L., Ball, L., & Cohen, D. K. (1996). Learning by all. *Phi Delta Kappan*, March, 468-476.

Wohlsletter, P. , & Briggs, K. L. (1994). The principal's role in school-based management. *Principal*, 74(2), 14-18.

### About the Author

**Rebecca Novick** is a research associate at the Northwest Regional Educational Laboratory's Child and Family Program and is currently conducting research in the area of developmentally appropriate practices, culturally responsive teaching, and professional development in early care and education. She has worked in early intervention as a classroom teacher, provided parenting education and support for parents involved with child protection agencies, and has experience in the areas of program development and evaluation.

Special interests include families at environmental risk and children in foster care. She holds a Ph.D in early childhood/special education from the University of Oregon. Correspondence may be sent to Rebecca Novick, 101 S.W. Main, Suite 500, Portland, OR 97204-3297 or via e-mail at novickr@nwrel.org.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

### Editorial Board

Greg Camilli
*camilli@rci.rutgers.edu*
John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Sherman Dorn
*sfxj9x@scfn.thpl.lib.fl.us*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**This article has been retrieved 481 times since September 6, 1996.**

**Education Policy Analysis Archives**

**Volume 4 Number 15** **September 6, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# A Review of Dorn's *Creating the Dropout*

**Sherman Dorn. (1996) *Creating the Dropout: An Institutional and Social History of School.* Praeger. $55.00.**

**Aimee Howley**
**Marshall University**

*ess016@marshall.wvnet.edu*

Let me recommend Sherman Dorn's new book, *Creating the Dropout*. The book undertakes a scholarly trek through the rhetoric of school leaving, construing economic and political vagaries as the occasions for a manufactured problem. At the end of the trip, the sympathetic reader is left wondering why he or she wasn't politically savvy enough back then to desert high school or, at the very least, to boycott the graduation ceremony.

Interesting as the historical journey proves, it somehow evades theoretical mapping, and this is a major weakness in an otherwise well-crafted effort. Throughout my reading, I kept taking side trips on my own to better situate Dorn's aims and interpretations. These provide a contrapuntal low road to the high one that Dorn has us travel.

Dorn begins his historical interpretation with a paradox: As increasing numbers of teenagers attended and graduated from high school, increasing rhetorical attention was drawn to the "dropout". This attention, however, took various forms at first, which crystallized into a set of predictable, stereotypic assertions in the 1960s. By the mid-1960s, in other words, graduation from high-school had become an age norm . But was failure to graduate really a crisis, either for the individual or for society? Or was its significance, its status as a "crisis" manufactured? In Dorn's view, the "drop-out" was invented, not discovered:

> ...dropping out in itself was not a primary concern of educators until the mid-twentieth century. Many of the issues we think of today as connecting with dropping out--the need to socialize children, the response of schools to urban poverty, the economic promise of education, and the problems of children who have academic difficulties in school--have appeared frequently without being part of an explicit discussion about dropping out. Only after 1960 did they become commonly identified as part of a specific problem called "dropping out." Concerns about dependency, the belief in schools' ability to improve the poor, and the expectation that all teenagers should be in school gelled in the dropout debate. Then educators struggled to respond to the "new" issue of dropping out. (p. 80)

The invention of the dropout was, according to Dorn, a way for schools and the media to channel and thus contain more general concerns about the condition of cities. Unlike the structural conditions of poverty or the irrelevance of the school curriculum, the dropout could be blamed for his (the invented dropout was most often male) own circumstances. Furthermore, he could be assigned blame for the increasing unrest within urban communities. In this manner, the effects of racism in the school and workplace, inadequate basic education, and unresponsive social services could be discounted. Schools and other government bodies could distance themselves, when the problems of the cities were attributed to some combination of inadequate upbringing, cultural disadvantage, and personal dereliction.

Dorn's explanation is compelling, and he supports it through a careful review of relevant professional literature about education as well as through an analysis of primary documents from three cities. Nevertheless, it is an interpretive claim, and its positioning as interpretation is not well enough explored. Because he avoids theoretical and methodological issues, Dorn leaves the reader to discover (or allows the reader to ignore) the sources of and supports for his underlying theoretical premise-- that discourse can invent social reality.

The tendency to draw this sort of conclusion has its own history, of course, and my first side trip was to find sources of this presumption. A cursory visit to the library catalog allowed me to identify an entire genre in historical and social science literature devoted to uncovering the social manufacture of certain real things that we all appear to take for granted, childhood, for example, (Aries, 1962), the "crisis of education" (Berliner, 1995), giftedness (Margolin, 1994), madness (Szasz, 1974). The analyses differ, but the leitmotifs are the same: the social world is something of our own making, not everything is what it seems. This approach to analysis, for which we might as well blame Marx (the hidden workings of the social relations of production) and Freud (the hidden psycho sexual motive) is itself an invention of discourse. Dorn, like the rest of us, is to some extent trapped in his own trap. In a world made of discourse, what truth claims can any discourse support? I found myself wishing that Dorn had wrestled more thoroughly with this fundamental question of purpose and method.

It would be unfair, however, to accuse Dorn of ignoring the question completely. He did deal with it in the context of his analysis of the rhetoric of "dropping out", but he construed it narrowly as if to imply that his own discourse and its moorings in a particular literature were somehow immune. His framing of the question looked something like this: Why was the social construction of the dropout crisis irrational? To understand what Dorn must mean by "rational," we can look at his answer:

> First, the perceived crisis was not in response to a real demographic trend; graduation became more, not less, prevalent in the middle twentieth century. Second, the perceived crisis did not lead to effective or even widespread policy changes. Third, the public debate over dropping out omitted issues and perspectives that a rational discussion should have included. (p. 99)

This answer suggests that a "rational" social construction would correspond to "the facts", support improvements, and attend to all the relevant issues. But isn't this asking too much of social construction? After all, the premise that something (the dropout, for instance) can be created out of the discourse surrounding it--in other words, can be interpreted into existence--suggests the presence, and in a logical sense, necessity, of multiple interpretations. If the facts manifested themselves apart from interpretation, we wouldn't need or, for that matter, even be able to tolerate discourse that subverted the self-evident "truth."

But facts, particularly about human enterprises, do not come to us that way. Nor do our interpretations, however earnest, require ameliorative action. Furthermore, interpretation, by its very nature, includes some and excludes other perspectives. In consideration of these features of interpretation, Dorn's invocation of the "rational" sounds antiquated and hollow. Rather than basing his claims on the impossible distinction between "rational" and "irrational" interpretations, Dorn would have been better served by examining the dynamics of conflict within the discourse itself. And to a certain extent--for example in his comparison of the Philadelphia school systems' claims about dropouts and the competing claims of a civil rights organization in West Philadelphia--he did. Nevertheless, this stance does not permeate the entire work. And, in my view, it should.

The most important side trip for me, then, involved reconstructing Dorn's argument in view of the assumption that the "dropout crisis"--by virtue of the fact that it could be nothing other than a social construct--was rational according to some logic. Finding the logic behind the construct became the purpose of my divagation. This low road came curiously close to the path that Dorn took in the final chapter of the book. But the divergences were also telling.

For Dorn, the dropout stereotype was important because of what it hid, not because of what it revealed. That is, by focusing on the dropout, educators and policy makers were able to shield themselves from direct confrontation with the inequities of schooling, the vagaries of the labor market, the paradoxes of credentialism, and the fear of dependency. This interpretation suggests that the particular construction of dropouts was intentional, rather than endemic. Educators, on this view, could have constructed matters otherwise. The "dropout" then hid from educators and the public an improved (liberal) prospect for education that might otherwise have been visible to them. In a broad sense, according to this interpretation, social construction is taken to be willful --the result of managed discourse, not of conflict over discourse.

The alternative reading, however, takes social construction to be the product of conflict whose sources arise outside of the discourse itself. On this view, social constructions embody material interests, and the conflicts over discursive representations of the social world implicate disputes over the way that material interests are translated into strategies of language. From this vantage, improvement has no absolute referent, and the truth of a claim depends on how it is contextualized, by whom, and toward what ends. This interpretation assumes that the position one takes on a question (for example, the question of dropouts) is not primarily voluntary, but constitutes an embodiment of one's material interests or alignments. Further, it posits that the truth of a claim is a matter internal to a position or constellation of interests, not susceptible to resolution across positions.

With respect to dropouts, the alternative reading presents two (or more) opposing sets of interests, reasoned in ways to establish internal coherence, but essentially incommensurable. One set of interests seeks to perpetuate social inequities, whether in the name of merit (e.g., recommending higher standards for degree attainment) or in the name of recuperation (e.g., calling for lower dropout rates). Providing more social goods to those who have historically been deprived constitutes another set of interests. And curiously, this set of interests may also be represented by the invocation to increase high school graduation rates of certain groups and to improve the quality of the high school curriculum.

Failing to give a thorough accounting of the conflicts implicit in the discourse on dropouts, Dorn ultimately provides a simplified and rootless interpretation. One of his concluding remarks demonstrates how this failing leads to a kind of incoherence.

> The way we have rationalized our expectation of graduation, with the stereotype of the high school dropout, has focused on the most superficial aspects of education--providing or maintaining the worth of credentials and preventing dependency and criminality. The social construction of the dropout problem has thus continued our national obsession with education either as a panacea for social problems or as the last bulwark against urban chaos. (p. 132)

What's wrong here is that Dorn imagines himself able to speak from some vantage external to social construction and, in a way, to discourse itself. If "we" are obsessed with a particular construction of education, how has Dorn managed to escape? If he hasn't escaped, how can he make the distinction between what is really "rational" and what is arbitrarily "rationalized?"

That this failing is subtle--some might say invisible or even manufactured--is testimony to Dorn's overall rigor and good will. He offers up a careful history in an effort to improve our outlook. The claim that his analysis of rhetoric might have opened onto a wider view of what discourse embeds and reveals is hardly a condemnation.

One last tangent took me back to the library for a brief and seemingly irrelevant, though surprisingly instructive, inquiry into the context of Dorn's title. I found him, and, for better or worse, he finds himself in the company of:

> *Creating the American Presidency, Creating the Best Impression, Creating the Big Game, Creating the Bill of Rights, Creating the Capacity for Attachment, Creating the Caring Congregation, Creating the Child, Creating the Child-centered Classroom, Creating the Cold War University, Creating the College of the Sea, Creating the Commonwealth, Creating the Competitive Edge through Human Resource Planning, Creating the Computer, Creating the Conditions for School Improvement, Creating the Constitution, Creating the Corporate Future, Creating the Country, Creating the Countryside, Creating the Couple, Creating the Empire of Reason, Creating the Entangling Alliance, Creating the Ergonomically Sound Workplace, Creating the European Community, Creating the Evangelizing Parish, Creating the Federal City, Creating the Federal Judicial System, Creating the Future, Creating the Future for South Dakota, Creating the Future of Health Care Education, Creating the Future Today, Creating the Future--Agendas for Tomorrow, Creating the Global Company, Creating the High Performance International Petrol, Creating the High Performance Team, Creating the Human Environment, Creating the Inclusive Preschool, Creating the Kingdom of Ends, Creating the Language of Thought, Creating the Library Identity, Creating the Literature Portfolio, Creating the Look, Creating the Medical Marketplace, Creating the Modern South, Creating the Multi-age Classroom, Creating the Nation in Provincial France, Creating the National Pastime, Creating the New American Hospital, Creating the New Local Government, Creating the New Wealth, Creating the Nonsexist Classroom, Creating the North American Landscape, Creating the Old Testament, Creating the Opportunity, Creating the Palestinian State, Creating the Peaceable School, Creating the People's University, Creating the Perfect Database, Creating the Perfect House Dog, Creating the Post-communist Order, Creating the Quality School, Creating the Resilient Organization, Creating the School, Creating the Second Cold War, Creating the Service Culture, Creating the Source through Folkloristic Fieldwork, Creating the Story, Creating the Successful Business Plan, Creating the Teachable Moment, Creating the Team, Creating the Technical Report, Creating the Technopolis, Creating the Thoughtful Classroom, Creating the Total Quality Effective School, Creating the Unipart Calendar, Creating the Virtual Store, Creating the Welfare State, Creating the West, Creating the Work you Love, Creating the World, Creating the Writing Portfolio, and Creating the 21st Century Through Innovation.*

Where exactly to locate Dorn's historical analysis among this crowd of persuaders, unpackers, and bandwagoneers is your decision. But despite a certain theoretical inattentiveness, he still occupies, in my view, a piece of the high ground.

## References

Aries, P. (1962). *Centuries of childhood: A social history of family life.* New York : Vintage Books.

Berliner, D.C., Biddle, B.J. (1995). *The Manufactured Crisis: Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison-Wesley.

Margolin, L. (1994). *Goodness personified: The emergence of gifted children.* New York: Aldine de Gruyter.

Szasz, T.S. (1974). *The myth of mental illness: Foundations of a theory of personal conduct* (rev. ed.). New York, Harper & Row.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@rci.rutgers.edu*
John Covaleskie
*jcovales@nmu.edu*
Andrew Coulson
*andrewco@ix.netcom.com*
Alan Davis
*adavis@castle.cudenver.edu*
Sherman Dorn
*sfxj9x@scfn.thpl.lib.fl.us*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Ernest R. House
*ernie.house@colorado.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
William Hunter
*hunter@acs.ucalgary.ca*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*
Dewayne Matthews
*dm@wiche.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Les McLean
*lmclean@oise.on.ca*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*
Richard C. Richardson
*richard.richardson@asu.edu*
Anthony G. Rud Jr.
*rud@purdue.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*
Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 16** **October 6, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of
Education, Arizona State University,Tempe AZ
85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# A Review of *Computers as Tutors: Solving the Crisis in Education*

**Frederick Bennett. (1996)** ***Computers as Tutors: Solving the Crisis in Education***

**Greg Sherman**
**Emporia State University**

*shermang@esumail.emporia.edu*

It was with great interest that I began reading Frederick Bennett's book *Computers as Tutors: Solving the Crisis in Education* (1996). Published on the Internet and located at http://www.cris.com/~Faben1/, Bennett's book not only represented the first complete book I have ever tried reading straight off the computer, but it also represented the only book on education I have ever read in which the title purported to have a solution to education's problems. It took me less than twenty minutes to discover that the book failed me miserably on both accounts.

I initially began reading *Computers as Tutors* by accessing the web site, skimming the prologue and table of contents, and then settling down in my office chair to commence reading and digesting Chapter One. With my hand on the mouse, I read the words and scrolled slowly down the Chapter One web page as needed. Things were going pretty well as I toggled between my web browser and a word processing program I was using to jot down notes. And then I began to realize that I wasn't paying close attention to the words. I was skimming and jumping up and down the page, scrolling to the bottom of the page to size up the chapter. I soon discovered that I was approaching this book the same way I approach most other web pages: skim the text, look for relevant information, and click on links that will take me to the precise information I desire. My brain was treating this on-line book like any other web site, and I couldn't concentrate. In addition, I couldn't get used to making notes on specific elements of the chapter by typing in a separate window. So I printed off the entire book; over 100 pages of single-spaced text. I three-hole punched the pages, put them into a binder, settled into my reading couch, and read. Much better.

Although *Computers as Tutors* was a rather lengthy read by web standards, the main points presented by Bennett were few and concise:

- Schools can use technology more effectively
- Schools must use technology differently
- Computers can remake education

- The key to utilizing computers more effectively is through their use as private tutors

Throughout the book, Bennett indicates emphatically that computers can solve most of the problems confronting educators today if computers are implemented as private tutors "...without a teacher interposed between the machine and the child." Bennett spends a good portion of the book describing all the specific benefits spawned by using the computer to provide effective, individualized instruction. These include relieving the teacher of burdensome and mundane teaching-related chores, providing an opportunity for all students to fulfill their need to succeed, accommodating the needs of the gifted and challenged students, reducing the need for substitute teachers, and eliminating prejudice against race and sex. In addition to these advantages, computer-based instruction could eliminate grades, promote better thinking skills, and provide a means of easily replicating and distributing successful learning programs. And because the use of computers has demonstrated the ability to improve reading skills, illiteracy could be wiped out, resulting in the reduction of such literacy-related problems as crime and poor job performance.

Before addressing what I feel are numerous flaws in Bennett's argument that computers as tutors can solve the problems facing educators today, I would like to point out some admirable strengths in the work. The writing itself is very well-structured, clear, and organized. Bennett describes many of the endemic problems within the institution of public education, and he identifies clearly the need for reform. Bennett astutely points out that computers are not being used to their potential and can play a vital role in a systemic reform movement. As they have done in the private and corporate sectors, better use of computers could provide greater flexibility in daily classroom scheduling, allow teachers to easily update and acquire effective materials, eliminate some paperwork, and accommodate absentees and nontraditional schedules.

There is no question that public education is in need of repair. There is no question that better use of computers can improve conditions in public education. And there is no question that students who perform well in school generally find themselves in better social and economic conditions when they emerge into the real world than students who perform poorly. Bennett does a commendable job of delineating the many ways computers can change how students might navigate through the system. But genuine reform isn't about changing how students learn. Genuine reform is more about changing what students learn, something Bennett's ideas regarding the use of computers in schools didn't even begin to address.

Near the beginning of the book, Bennett states: "When American education fully embraces computerized education, the dreadful state of American schooling will change overnight. Almost every child in the United States will learn to read early in their schooling. They will then be able to enjoy education." The implications of this statement are twofold: 1) the key to success in education is literacy, and 2) traditional, text-based instruction should be perpetuated. Bennett supports his literacy approach to reform by indicating that people who participate in riots, commit felonies, have out-of-wedlock births, or depend on welfare for support are more illiterate than people who don't exhibit such behaviors. Reduce the number of illiterate people, Bennett argues, and these types of behaviors will decrease. It is certainly beyond the scope of this review to speculate on whether or not reducing illiteracy will reduce poor decision-making, but my gut tells me that, like crime and welfare dependency, illiteracy is probably a symptom of a much bigger societal problem.

Bennett places a high educational premium on literacy, and he maintains that computers as individual tutors can get students reading better, faster, sooner. He contends that computers haven't had much of an impact in education because they have not been used as teachers. "This failure to allow computers to teach is the reason technology thus far has been a dismal failure in schools." He uses examples of how individual tutors have had profound impacts on the lives of successful people such as Alexander the Great, John F. Kennedy, and Thomas Edison. He describes how Edison was removed from school at an early age, yet excelled as a result of individual instruction from his mother. I certainly agree that Edison's mother probably had a positive influence on his development as a creative inventor, but I am quite certain his achievements were not the result of the effective instruction of school-related educational outcomes. People don't learn to become great inventors because somebody taught them to read. People become great inventors because somebody taught them to be great inventors. Edison's early education was probably more about exploration and intellectual encouragement than it was about reading.

Reading may have played a part in Edison's early education, but it was certainly not the goal of his education. Referring to his mother as tutor, Edison said "She instilled in me the love and purpose of learning." Implicit in this statement is the purpose of true educational reform: change WHAT is taught, not how. Bennett's book actually encourages the status quo in this area. For example, Bennett states that "...computerized education will be far more efficacious for developing better reasoning skills." He then describes what he feels are the three requirements for developing better reasoning skills: good underlying education, thought provoking questions, and time to respond to these questions. Based on his ideas up to this point, we can only assume that "good underlying education" refers in no small part to literacy. And "thought-provoking questions" still places this type of educational experience in the realm of text-based instruction. Not to mention that this Aristotelian pedagogical approach represents a rather simplistic formula for developing higher-order thinking skills. If it were this easy, there would be very little need for any technology in the learning process. What Bennett fails to address are the opportunities to use computer-based technology as contexts for experiencing purposeful, meaningful instructional environments where learning to read, performing mathematical calculations, and operating at higher levels of reasoning are not the end of the instruction but the means to a purposeful end.

If they are to be used effectively, computers should be part of an instructional environment which supports the learning of skills that students will need in order to be successful in the real world. Reading may be a prerequisite for many of these real-world outcomes, but believing the computer can successfully deal with all the outcomes related to literacy, including choosing to read, is narrow and misguided. Bennett states: "[Computers] can communicate information more efficiently and they can do it with a certain panache-they can fascinate while they teach." Substitute the word "television" for the word computers and you echo the sentiment of educational reformers in the 1950's who believed technology was really going to have an impact on how students learned. And like any other piece of instructional hardware, computers probably won't have a profound impact on how anybody learns anything. Somebody may be able to learn how to read from a computer as tutor because they have an opportunity to practice practice practice, with a certain level of feedback provided. But in the end, this is no different than working with an individual or a small group. The computer may be able to facilitate learning to read in a more efficient manner, but this is no indication that the learner will choose to read outside school, or will choose to communicate in written form, or will enjoy any or all of it.

But like television, computers can make a difference in what is learned. Because of television, many people in the United States have learned that owning lots of different, new products is important. As a "window to the world" television has also helped us to know more about people from other countries, and good or bad we know that reading isn't the only way to obtain information about the world around us. Because of computers, we can easily communicate in writing to people all around the world, we can access precise information needed in a number of ways, we must discern between the relevant and the irrelevant, and we can create, simulate, and explore in countless ways. These are the reasons why computers can make a difference in schools. These indicate that different things can be learned in school. Like Edison's mother, computers can be used to provide a purpose for learning things that are important to us. And these types of outcomes go far beyond and around literacy.

Bennett summarizes his work by stating that "Computerized education will mean a profound alteration in the manner in which schooling is carried on." Bennett does a good job of pointing out exactly how schooling could change as a result of using computers as tutors. But no reform movement is carried very far by addressing schooling. We need to address learning, which isn't necessarily related to schooling. So if you want to read about all the different ways computers can address more effective ways of doing what public education tries to do today, read Frederick Bennett's Computers as Tutors: Solving the Crisis in Education. But if you think the crisis in education has something to do with what education tries to do today, you would be better off reading Seymour Papert's *The Children's Machine* or Howard Gardner's *The Unschooled Mind*. These books address real change and real reform. And although you can't access them on the Internet, you will probably save in the long run because they are already printed out for you.

## About the Author

Greg Sherman is an Assistant Professor in the Division of Instructional Design and Technology, The Teachers College of Emporia State University

shermang@esumail.emporia.edu

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

### Editorial Board

Greg Camilli
*camilli@rci.rutgers.edu*

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Sherman Dorn
*sfxj9x@scfn.thpl.lib.fl.us*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 17 November 11, 1996 ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDU POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# What Does the Psychometrician's Classroom Look Like?: Reframing Assessment Concepts in the Context of Learning

**Catherine S. Taylor**
**University of Washington**

*ctaylor@u.washington.edu*

**Susan Bobbitt Nolen**
**University of Washington**

*sunolen@u.washington.edu*

## Abstract

We question the utility of traditional conceptualizations of validity and reliability, developed in the context of large scale, external testing, and the psychology of individual differences, for the context of the classroom. We compare traditional views of validity and reliability to alternate frameworks that situate these constructs in teachers' work in classrooms. We describe how we used these frameworks to design an assessment course for preservice teachers, and present data that suggest students in the redesigned course not only saw the course as more valuable in their work as teachers, but developed deeper understandings of validity and reliability than did their counterparts in a traditional tests and measurement course. We close by discussing the implications of these data for the teaching of assessment, and for the use and interpretation of classroom assessment data for purposes of local and state accountability.

More than ever before, pressure is being placed on teachers to create high quality assessments of their students' learning. Work is underway in Kentucky, New Mexico, Vermont, Washington, and in the eighteen states that are members of the New Standards Project (Resnick and Resnick, 1991) to explore the viability of classroom-based assessments, projects, and portfolios as sources of state or national accountability data about student learning. These initiatives emerge from a growing belief that the teacher can be one of the best sources of information about student learning. However, there is growing evidence that teachers have not been adequately prepared to create and conduct valid assessments. Even teacher education programs that include an assessment course may not help teachers develop the concepts and skills necessary to meet these assessment demands.

To address this problem, districts, states, and national organizations have invested considerable resources in in-service training for teachers. Organizations such as the National Council on Measurement in Education (NCME) and the Association for Curriculum Development and Supervision (ASCD) have developed training modules and training materials for classroom teachers. Groups such as the National Council for Teachers of Mathematics (NCTM) have developed documents such as *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions* (NCTM, 1991) and *Assessment Standards for School Mathematics Standards* (NCTM, 1995) in an attempt to help teachers incorporate more appropriate assessments into their teaching practices. Still, these efforts may not be successful if the models used to educate teachers in the concepts and skills of assessment do not fit the reality of classrooms.

An example of the confusion caused by the mismatch between models based on test theory and the demands of the classroom context illustrates this problem. Preservice teachers in an assessment class had read Smith's (1991) article on the meanings of test preparation. Smith lists a number of ways teachers prepare for external standardized tests, including teaching the specific content covered on the test. Students were surprised to find that psychometricians considered this to be cheating. Were they not being admonished, by both the instructor and the course textbook to do just that--assess to see whether students were learning what had been taught. To these students, if vocabulary words were to be tested, they should be taught. If science or social studies concepts and facts were to be tested, they should be taught. Even if the test expected students to generalize a concept or skill to the new situation, the concept or skill should have been taught first! In the words of one puzzled student, "What does the psychometrician's classroom look like?"

This apparent discrepancy between the idea of "domain sampling" central to test theory and the notion that classroom assessment is intended to assess whether students learn what they are taught arises from a clash of contexts. The world of large scale external tests is very different from the world of the classroom. In this paper, we will argue that traditional tests and measurement courses and most assessment textbooks for teachers present measurement concepts in ways that better fit the world of external tests designed to measure individual differences. When teachers are taught traditional measurement concepts and expected to apply them to the context of teaching and learning, they have little chance of developing the skills and concepts they need to assess their students. We will also argue that the meanings of assessment in the context of the classroom must be considered carefully when large scale assessment programs decide to use classroom assessments for the purposes of district, state, or national accountability.

We begin by challenging traditional notions of testing and measurement in terms of their fit to the classroom. While we recognize that the principles of classical test theory may be appropriate for some contexts (e.g., administering and interpreting standardized, norm-referenced tests), we see a need for more clarity in how these models, their applications and limitations, are presented to teachers. We discuss the theoretical underpinnings of traditional measurement concepts and why they must be reframed in light of the classroom context. We examine the ways in which reliability and validity are presented in eight recently published assessment texts designed for teacher preparation and discuss why definitions of validity and reliability presented in most educational assessment textbooks fit the context of external testing better than that of the classroom.

Next we present frameworks for validity and reliability that situate these constructs in the world of the classroom teacher, and discuss how these frameworks might be used in teacher education. We then present an overview of the assessment course we developed to help preservice teachers understand the concepts of validity and reliability as they are reframed in this paper. The work of the course was designed to help preservice teachers develop a deep understanding of the potential relationship between classroom assessment practices, subject-area disciplines, and instructional methods so that they would see valid and reliable assessment as central to their work as teachers. Evidence for the effectiveness of basing our assessment course on these frameworks is provided in the form of three studies comparing the responses of students in the redesigned course to those taking a traditional tests and measurement course in the same teacher education program.

We discuss the need for the measurement community to acknowledge the differences between the methods appropriate for external measurements and the measurement of the learning targeted by classrooms and schools. We suggest that those who prepare assessment text-books for the preparation of teachers, as well as instructors of assessment courses, clarify the philosophical positions underlying different assessment purposes and present assessment concepts in ways that are consistent with those differing purposes rather than attempting to blend frameworks that come from different philosophies about the purposes of assessment. Finally we discuss what these classroom-based conceptions of reliability and validity suggest in terms of what constitutes appropriate classroom-based evidence for large scale assessment programs.

## The Misfit of the Measurement Paradigm

The classroom context is one of fairly constant formal and informal assessment (Airasian, 1993; Stiggins, Faires- Conklin, & Bridgeford, 1986). However, few teacher preparation programs provide adequate training for the wide array of assessment strategies used by teachers (Schafer & Lissitz, 1987, Stiggins & Bridgeford, 1988). Further teachers do not perceive the information learned in traditional tests and measurement courses to be relevant to their tasks as classroom teachers (Gullickson, 1993; Schafer & Lissitz, 1987; Stiggins & Faires-Conklin, 1988). Wise, Lukin, and Roos (1991) found that teachers do not believe they have the training needed to meet the demands of classroom assessment. At the same time, teachers' ability to develop appropriate classroom-based assessments is seen as one of the six core functions of teachers (Gullickson, 1986).

Several authors have outlined what they believe are the essential understandings about assessment teachers must have in order to confront the ongoing assessment demands in the typical classroom (Airasian, 1991; Linn, 1990; Schafer, 1991; Stiggins, 1991). Many of these concepts and skills, as well as those presented in measurement text-books (e.g., Hanna, 1993; Linn & Gronlund, 1995; Mehrens & Lehmann, 1991; Nitko, 1996; Oosterhof, 1996; Salvia & Ysseldyke, 1995; Worthen, Borg, & White, 1993), are derived from a model of measurement that began in the late 1800s. Rooted in scientific thinking of the nineteenth century, test theory is based on a model of the scientific method.

With classroom instruction as the equivalent of a treatment, test theory would suggest that tools of assessment are designed to carefully assess the success of instruction for different examinees. Taking the perspective of Galton (1889), students differ in their inherent capacity to learn the content of various disciplines. The assessor is the scientist who must dispassionately assess and record each students' attainment of the defined outcomes of instruction. Students are the focus of observation and the measurement model presumes them to behave like passive objects. As Cronbach (1970) noted,

> A distinction between standardized and unstandardized procedures grew up in the early days of testing. Every laboratory in those days had its own method of measuring. . . and it was difficult to compare results from different laboratories. . . Standardization attempts to overcome these problems. A standardized test is one in which the procedure, apparatus, and scoring have been fixed so that precisely the same testing procedures can be followed at different times and places. . . If standardization of the test is fully effective, a man will earn very nearly the same score no matter who tests him or where. (pp. 26-27, italics added)

The classroom teacher, however, is not a dispassionate observer of students' learning. Classroom teachers have a vested interest in the outcomes of instruction--many believing that student failure is a reflection on their teaching. Both the popular press and current legislation in states such as Kentucky would suggest that the public agrees with this view of the relationship between teaching and learning. The classroom teacher, in contrast to the experimental scientist, is more like a "participant observer" (Whyte, 1943). Using the words of Vidich and Lyman (1994), the teacher is much like an ethnographic researcher. In the following quote, the authors' use of the term "ethnographic researcher" has been replaced by the term "teacher."

> The [teacher] enters the world from which he or she is methodologically required to have become detached and displaced. . . . [T]his [teacher] begins work as a self-defined newcomer to the habitat and life world of his or her [students]. He or she is a citizen-scholar as well as a participant observer. (Vidich & Lyman, 1994, p. 41)

Teachers adjust instruction for the needs of students; adapt instruction for the needs of diverse students; bring a wide range of evidence to bear on decision-making about students - extending beyond the evidence from standardized tests to observations of students' classroom behaviors, attitudes, interests, and motivations (Airasian, 1994). The purpose of classroom assessment is to find out whether students have benefited from instruction. However, unlike the dispassionate observer, the good teacher regularly adjusts the treatment, in response to ongoing assessments, in order for learning to be successful.

While the participant observer may be required to use certain methods to increase their "objectivity," they must both observe and participate in the world of the classroom. They "make their observations within a mediated framework, that is, a framework of symbols and cultural meanings given to them by those aspects of their life histories that they bring to the observational setting" (Vidich & Lyman, 1994, p. 24). The teacher's decision to attend to one source of assessment information over another reveals as much about the "value-laden interests" of the teacher as it does about the subject of her/his assessments (Vidich & Lyman, 1994, p. 25).

While this may be seen by measurement professionals as the reason objective measures are needed, qualitative researchers would respond that "The more you function as a member of the everyday world of the researched, the more you risk losing the eye of the uninvolved outsider; yet, the more you participate, *the greater your opportunity to learn.*" (Glesne & Peshkin, 1992, p. 40, italics added). Qualitative researchers would say that the very choice of what items to include in a test reflects the values and biases of the teacher. Hence the job of those who prepare teachers for classroom assessment must include an awareness of the context in which teachers teach, the goals of instruction and schooling, and the complex demands of the work of a participant observer.

If teachers are not dispassionate observers, neither are students passive objects. They are influenced by assessment processes and products (Bricklin & Bricklin, 1967; Butler, 1987; Covington & Beery, 1976; Deci & Ryan, 1987). They adapt their approach to learning and preparation for assessment in order to gain the highest possible scores (Toom, 1993). They may take on persona that will afford them the grace of teachers. Hence, neither teachers nor students fit the scientific model of standardized measurement used to frame the measurement concepts and strategies taught to teachers.

## Assessment and Teacher Preparation Programs

Despite the importance of assessment in the experience of students and in teachers' ability to determine the success of instruction in terms of student learning, assessment instruction is peripheral in many teacher education programs. In programs that do include assessment courses, assessment is usually treated as a foundational course focused on a set of generalizable concepts and skills. In most programs, all prospective teachers, from the kindergarten teacher, to the APP calculus teacher, to the middle school vocal music teacher are taught in a single group. In others, assessment instruction is relegated to a 1-2 week unit in an omnibus educational psychology course. In response to the formidable range of assessment content teachers need to know, instructors may design courses that result in intellectual awareness of key concepts rather than actual competency in applying. Research on the professional development of teachers (e.g., Cohen & Ball, 1990; Grossman, 1991) suggests that intellectual awareness is not sufficient to overcome the "apprenticeship of observation" (Lortie, 1975) that dominates pre-service teachers' learning. Without significant intervention, pre-service teachers typically adopt the practices that were used with them as students or those that are used by their cooperating teachers.

Assessment textbooks generally reflect a view of assessment courses as survey courses, intended to present a range of assessment ideas and leaving to instructors (or the students themselves) the task of constructing a coherent picture of assessment. As Anderson, et al (1995) have noted, survey approaches to the preparation of teachers do not allow for a "rich and grounded" understanding. Ironically, textbook authors' attempts to acknowledge the classroom context may contribute to teachers' confusion and antipathy. Many textbooks (e.g., Hanna, 1993; Linn & Gronlund, 1995; Mehrens & Lehmann, 1991; Oosterhof, 1996; Salvia & Ysseldyke, 1995; Worthen, Borg, & White, 1993) combine presentations of assessment in the classroom with traditional presentations of the principles of testing and basic concepts of measurement. As we will argue in the next section, the notions of validity and reliability used in large scale external testing must be recast before they can be useful in the context of classroom teaching and learning. With the increased emphasis on appropriate assessment practices in the classroom, we must take seriously the gulf between what classroom teachers believe they need to know about assessment and what measurement professionals believe teachers need to know. In the next sections, we provide frameworks for bridging this gulf.

## Definitions of Validity

### Traditional Presentations of Validity

All of the assessment text books reviewed for this article acknowledged the contextual issues in the classroom; however, chapters on validity generally used the language of scientific methodology to describe this construct. Most of these texts (e.g., Hanna, 1993; Linn & Gronlund, 1995; Nitko, 1996; Salvia & Ysseldyke, 1995; Oosterhof, 1996; Worthen, Borg, & White, 1993) presented three or four "types" of validity: construct validity, content validity, criterion- related (predictive and/or concurrent) validity, and recommend that evidence for each type of validity should be obtained when using a test. Measurement professionals generally agree that for assessments to be valid, they should (a) measure the construct they are intended to measure, (b) measure the content taught, (c) predict students' performance on subsequent assessments, and (d) provide information that is consistent with other, related sources of information. Consequences of test interpretation and use, a validity issue recently raised by Messick (1989), is addressed by few published classroom assessment texts (For example, see Hanna, 1993; Linn & Gronlund, 1995; Nitko, 1996). In fact, some would disagree that "consequential validity" is a component of the construct of validity at all (See Stuck, 1995).

Traditional presentations of these types of validity often define evidence for validity in terms of: (a) correlations between tests measuring the same construct or between a test and the criterion behavior of interest (Hanna, 1993; Linn & Gronlund, 1995; Nitko, 1996; Worthen, Borg, & White, 1993), (b) tables of specification to determine whether the content of a test measures the breadth of content targeted (Linn & Gronlund, 1995; Mehrens & Lehmann, 1991; Oosterhof, 1996), and (c) using a range of strategies to build a logical case for the relationship between scores from the assessment and the construct the assessment is intended to measure (Linn & Gronlund, 1995; Nitko, 1996; Oosterhof, 1996).

These types of validity evidence are based on two different notions of what makes an assessment valid. The evidence for the validity of an assessment is provided if (a) students perform consistently across different measures of the same construct (a notion that comes from a theory of individual differences (Galton, 1889)) and (b) links between what is measured and some framework or context external to the test (Linn & Gronlund, 1995; Messick, 1989). Taken individually, these two prongs of validity theory do not have equal value in the classroom. Classroom teachers are less interested in the consistency of student performance across similar measures than they are in whether students' learn what they are teaching (the targeted constructs). Learning, especially of skills and strategies that are taught throughout schooling, is expected to change rather than remain consistent over time.

Consistency with other, related performances is also problematic for teachers as they teach each new group of students. Given the option of looking over prior school records, teachers often claim that they do not want to be prejudiced by others' views (Airasian, 1991, p. 54). Over the course of a year, inconsistent performance may be attributed to many factors other than the validity of assessments. Students who begin to perform more poorly than expected may be informally assessed through interviews with the students and reviews of their work. Teachers may become alarmed and contact school support staff and/or parents to see if the cause lies outside the classroom. On the other hand, when poorly performing students begin to dramatically improve performance, teachers may see this as evidence of student learning and of their own success as teachers. Consistent performance across assessments is only desirable when performance is consistently good or when the content taught is constantly changing (e.g., spelling lists).

As Moss's (1996) paper suggests, the notion of the assessor as "objective observer" does not fit the context of the educational assessment as well as it does the work of experimental science. Teachers see students as the focus of purposeful action (Bloom, Madaus, & Hastings, 1981). Tests and other assessments provide information, not only about how well students have learned, but about how well they are presenting the targeted content and concepts (Airasian, 1993; Mehrens & Lehmann, 1991; Nitko, 1996; Oosterhof, 1996), how students are feeling about school, themselves, and their worlds (Airasian, 1993). Hence it is the responsibility of measurement professionals to help teachers learn how to choose and create assessment tools that will do the best job possible to make appropriate decisions about students' learning. This requires teachers to have a clear notion of validity that fits the work and the world view of teachers.

### Validity in the Classroom Context

In this section, we situate Messick's (1989) dimensions of validity in the context of classroom teachers' decision-making. Messick claimed that construct validity is the core issue in assessment, and stated that all inferences based upon, and uses of, assessment information require evidence that supports the inferences drawn between test performance and the construct an assessment is intended to measure.

> We can look at the content of the test in relation to the content of the domain of reference. We can probe the ways in which individuals respond to the items or tasks. We can examine the relationships among responses to the tasks, items, or parts of the test, that is, the internal structure of test responses. We can survey relationships of test scores with other measures and background variables, that is, the test's external structure. We can investigate differences in these test processes and structures over time, across groups and settings, and in response to . . . interventions such as instructional . . . treatment and manipulation of content, task requirements, or motivational conditions. Finally, we can trace the social consequences of interpreting and using test scores in particular ways, scrutinizing not only the intended outcomes, but also the unintended side effects. (Messick, 1989, p. 16)

Validity, then, is a multidimensional construct that resides, not in tests, but in the relationship between any assessment and its context (including the instructional practices and the examinee), the construct it is to measure, and the consequences of its interpretation and use. Translated to the classroom, this means that validity encompasses (a) how assessments draw out the learning, (b) how assessments fit with the educational context and instructional strategies used, and (c) what occurs as a result of assessments including the full range of outcomes from feedback, grading, and placement, to students' self concepts and behaviors, to students' constructions about the subject disciplines.

Messick stated that multiple sources of evidence are needed to investigate the validity of assessments. In the classroom context, this means that teachers must know how to look at their own assessments and assessment plans for evidence of their validity, they must know where to look for alternative explanations of student performance, and they must consider the consequences of assessment choices on their students and themselves. In short, teachers should develop a "habit of mind" related to their assessment processes. After situating each dimension in the context of teachers' work, we suggest general approaches that assessment instructor might use to help teachers use that dimension in their own assessment practice.

*Validity Dimension 1: Looking at the content of the assessment in relation to the content of the domain of reference.* Before teachers can look at their assessments in this way, they must be able to think clearly about their disciplines, understanding both the substantive structure (critical knowledge and concepts) and the syntactic structure (essential processes) of the disciplines they teach (Schwab, 1978). They must be able to determine which concepts and processes are most important and which are least important in order to adequately reflect the breadth and depth of the discipline in their teaching and assessments. As Messick (1989) states, one of the greatest sources of construct invalidity is over- or under-representation of some dimension of the construct. Once they have clearly conceptualized the disciplines they teach, teachers must know how to ascertain the degree to which the types of assessment tasks used in the classroom are representative of the range and relative importance of the concepts, skills, and thinking characteristic of subject disciplines.

In addition, because the process of assessment is as much a function of how assessments are scored as it is a function of whether the tasks elicit student learning related to the structure of the discipline, teachers must examine the degree to which the rules for scoring assessments and strategies for summarizing grades reflect the targeted learnings. As with breadth and depth of coverage within assessments, teachers must be able to evaluate whether scoring rules give too little or too much value to certain skills, concepts, and knowledge leading to questions about the validity of the interpretations teachers make from resulting scores.

To obtain evidence for this dimension of validity, teachers can be taught to stand back from their teaching, frame the learning targets of instruction carefully, and plan instruction and assessment together, in light of the overall targets of instruction. Without a clear picture of what is to be accomplished in a course or subject area, teachers cannot adequately assess whether their assessments (selected or self-developed) are valid. Once teachers develop a framework of learning targets (learning goals and objectives), they can learn how to carefully analyze whether assessment and instructional decisions link back to this framework. They can be given opportunities to look at scoring rules developed for open-ended student work and determine whether these rules relate directly to these targets of learning.

*Validity Dimension 2: Probing the ways in which individuals respond to the items or tasks and examining the relationships among responses to the tasks and items.* Teachers do not often have the luxury of "item tryouts" when developing their assessments. Before giving students an assessment, teachers must examine the degree to which the assessments have the potential to elicit the learning the students are expected to achieve. This means they must examine the assessment tasks and task directions to determine whether students are really being asked to show the learning related to the targets. Teachers must know to ask themselves, "Have the directions for the task or the wording of the items limited my students' understanding of the expectations of the task?"

Teachers should be encouraged to use assessment strategies that will allow them to probe their students' thinking and processes. This becomes increasingly important as the emphasis on higher level thinking and processes increases (Stiggins, Griswold, & Wikelund, 1989). In performance assessments, for example, examinees are often asked to explain their thinking and reasoning as part of the assessment task. Teachers commonly ask students to show their work in mathematics and science assessments. These classroom assessment practices lend themselves to probing the ways in which individuals are responding. This probing not only provides information about the validity of the assessments, but can provide better pictures of students' learning.

Teachers must know how to look across students' responses to a variety of assessment tasks to determine whether patterns of students' responses support the use of the assessments. The mechanisms for this type of examination have historically been quantitative item analysis techniques. However, few teachers use these quantitative techniques in actual classroom practice (Stiggins & Faires-Conklin, 1988). Teachers can be shown how to scrutinize student work qualitatively, looking for patterns in responses that reveal positive and negative information about the assessments. If items and tasks have not yet been used with students, teachers must know how to examine the demands of a range of items and tasks and ask themselves, "Are students who can show understanding of a concept in one assessment format (e.g., an essay), likely to show equal understanding in a different format (e.g., a multiple-choice test)?"

In order to probe examinee performance within and across different measures, teachers can learn to develop multiple measures of the same targeted learning. They may not only discover different ways to assess a given construct, but they may discover for themselves that particular types of assessment are more or less suited to certain learning targets.

*Validity Dimension 3: Investigating differences in assessment processes and structures over time, across groups and settings, in response to instructional interventions.* To investigate these validity issues, teachers must know how to examine the relationship between the instructional practices used and the assessments themselves. They must ask themselves, "Did I or will I actually teach these concepts well enough for students to perform well?" They must also evaluate the adequacy of various assessment strategies for the unique needs of their students. They must be able to judge whether an assessment can be used in many different contexts or whether differing contexts, groups, and instructional strategies require the development of different assessments.

Examination of this dimension of validity can be obtained when teachers are asked to look carefully at the relationship between an instructional plan and the demands of an assessment. If the work demanded in an assessment was not an adequate focus of instruction, teachers can decide ahead of time whether to adjust instruction to fit the learning targeted in the assessment or whether to adjust assessments to fit the learning targeted in the instruction.

*Validity Dimension 4: Surveying relationships between assessments and other measures or background variables.* Teachers must know how to judge the degree to which performance on the assessment and the score resulting from the assessment are directly attributable to the targeted learning. They must determine whether performance is influenced by factors irrelevant to the targeted learning such as assessment format, response mode, gender, or language of origin. This becomes increasingly critical as classrooms become more diverse and whole group teaching becomes more difficult. In general terms, teachers must know how to adapt an assessment format to meet the needs of diverse students while still obtaining good evidence about student learning related to the targets of instruction. Finally, teachers must know how to create scoring mechanisms for open-ended performances that are clearly related to the learning targets and that are precise enough to prevent biased scoring.

When teachers develop assessments, they can be asked to examine whether factors other than the targeted learning will influence students' performances. They can be asked to examine scoring rules to see whether the rules provide an unfair advantage or disadvantage to students who have certain strengths or weaknesses unrelated to the targeted learning.

*Validity Dimension 5: Tracing the social consequences of interpreting and using test scores in particular ways, scrutinizing not only the intended outcomes, but also the unintended side effects.* Teachers must consider the influence of classroom assessments on the learners themselves. The nature of the assessments, feedback, and grading can all influence student learning, students' self concepts and motivation (Butler & Nisan, 1986; Covington & Omelich, 1984), and their perceptions of the disciplines being taught. Teachers who assess their students' knowledge of science by giving them only multiple-choice tests of isolated facts, for example, communicate that science is a collection of facts about which everyone agrees. Those who assess students' inquiry strategies and their ability to make generalizations from observations or to systematically test their own hypotheses, communicate something different about the structure of the discipline of science.

To examine this dimension of validity, teachers can be asked to assess whether a given assessment reflects the syntactic and/or substantive structure of the discipline they teach (Schwab, 1978). Does the assessment target students' deep understanding of important concepts within the discipline or does it test surface knowledge? Does the assessment ask students to show their ability to use the processes through which professionals within the discipline construct new knowledge and ideas?

Teacher also can be asked to determine whether methods used to summarize grades for a marking period give adequate weight to those performances most directly related to the learning targeted. Teachers can be asked to look at their methods of feedback (formative assessments) and determine whether they are likely to motivate learning or to stifle learning; to assess whether feedback will lead to improvement, be largely insubstantial (Sommers, 1991), or be perceived by students as too late to make a difference in their grades (Canady, & Hotchkiss, 1989).

The five dimensions of validity described here can be taught in ways that emphasize their importance and usefulness in teachers' everyday work. Later we will briefly describe a course designed for this purpose and present evidence of its effectiveness. We recognize, however, that validity rests, in part, on teachers' ability to gather reliable information about student learning. Traditional presentations of reliability, based on test theory, are not immediately transferable to the work teachers do. In the next section, we describe traditional treatments of reliability in assessment textbooks and present an alternative framework.

## Dimensions of Reliability

### Traditional Presentations of Reliability

Measurement professionals place most of their emphasis in assessment on reliability--often at the expense of the validity of assessments. A common claim in test theory is that "for an inference from a test to be valid, or truthful, the test first must be reliable." (Mehrens & Lehmann, 1991, p. 265). This assumption is based on a mathematical model of test theory wherein observed scores are composed of true scores and measurement error. The less error in a test (i.e., the more reliable) the more truthful the test score. Hence, an unreliable assessment is automatically less valid.

Textbooks usually discuss reliability in terms of consistency (Airasian, 1993; Hanna, 1993; Linn & Gronlund, 1995; Mehrens & Lehmann, 1991; Nitko, 1996; Oosterhof, 1996; Salvia & Ysseldyke, 1995; Worthen, Borg, & White, 1993). When gathering evidence for the reliability of tests, the focus on consistency is related to either score reliability or rater reliability. Score reliability means that if a test were administered to an examinee a second time, the examinee would receive the same or about the same score. One way that measurement specialists try to ensure score reliability is through the standardization of tests. When assessments are standardized, all examinees complete the same items and/or tasks. If examinees are retested, they should complete the exact same tasks under exactly the same conditions. This would help to ensure that consistency of performance.

Another element of score reliability discussed in textbooks is that of generalizability. The longer the test (the more items and tasks) the more opportunities students have to show their learning. If students do better than they should on one item or task, they are just as likely to do more poorly than they should on another item or task. If a test is long enough, positive measurement error should cancel negative measurement error. Hence, the student is likely to earn a score that would be replicated if s/he took a parallel test. Writers who have expanded their discussion of reliability to include performance-based assessments focus on the number of performances necessary to obtain scores for examinees that can be generalized to the domain of interest (Linn & Burton, 1994).

Discussions of reliability in many textbooks; however, are based on the notion that assessment takes place at a single time and that summary decisions are made about examinees based on single testing events. In the classroom, teachers are engaged in on-going assessment over time and across many dimensions of behavior (Airasian, 1993; Stiggins, Faires-Conklin, & Bridgeford, 1986). Like motivation researchers, teachers see giving students choices about assignments as a way to increase student motivation and engagement (Deci & Ryan, 1985; Nicholls, 1989; Nicholls & Nolen, 1993). While individualization of instruction may result in better achievement and motivation, it means that standardization is very difficult. In addition, few teachers have the time or the inclination to administer parallel test forms to see whether students' scores are consistent; and psychometric techniques developed for looking at internal consistency of exams are not appropriate for many forms of classroom assessment. Some teachers give students opportunities to revise their work after feedback, both for the purposes of assessment and to enhance student learning (Wolf, 1991). Hence, the notion of a test with multiple items is only one of many possible assessment episodes in the classroom. Teachers do, however, collect many sources of information about student learning--not only through tests but through a range of formal and informal assessments: homework, classroom work, projects, quizzes. If this information is relevant to their learning targets, teachers could make reasonable generalizations about student learning.

The second dimension of reliability relates to the judgments made about students' work. Rater reliability refers the degree to which raters agree when assessing a given student's work.

Studies have documented that when raters are well trained and scoring criteria are well developed, raters can score student work with a high degree of consistency across raters (e.g., Hieronymus & Hoover, 1987; Shavelson & Baxter, 1992). In the classroom, however, a single judge (the teacher or a teaching assistant) is often responsible for evaluating all student work. Teachers rarely exchange student work or have another evaluator look at student work.

### Reliability in the Classroom Context

For reliability to have meaning for teachers, the concept has to make sense for the classroom and school context. Two dimensions of reliability relevant to the classroom are:

*Reliability Dimension 1: Determining the dependability of assessments made about students.* The concept of reliability can be reframed to fit the classroom context if the reality of the classroom and a broader and inclusive meaning of reliability are acknowledged. The American Heritage Dictionary (Houghton Mifflin Company, 1981) definition of reliable is "dependable." While measurement professionals have equated dependable with consistent, the former term is more appropriate for the classroom. Assessment may occur frequently in the classroom using measures that could not stand up to psychometric standards of reliability (e.g., research reports, written essays); however, it is possible that grading decisions made at the end of a marking period can be much more reliable than the individual assessments themselves. Even writers who are fairly cautious about performance-based assessments and portfolios admit that the classroom context could provide more reliable assessment information simply because teachers have more information from which to make judgments (Dunbar, Koretz, & Hoover, 1991). Hence, for assessments to be reliable, teachers must ensure that they have sufficient information from which to make dependable decisions about students. Given this framework, evidence for the validity of assessments used to make decisions should be the foremost consideration for teachers. Reliability of assessment decisions depends on the quality of the assessments. If attention is given to evidence for validity, then teachers can begin to ask themselves whether there is sufficient information from which to make dependable decisions. A wide range of assessments can serve the purpose of a long test--the more sources of assessment information, with demonstrable evidence for validity, the more likely dependable decisions can be made.

Teachers can be asked to look across diverse sources of assessment information planned for a given unit of instruction and determine whether there is sufficient information from which to make dependable judgments about students' learning related to the learning targets for the unit. Teachers can use grading policies to organize their thinking about the sources of information available for making judgments about student learning. Rather than using "averaging" techniques in grading, teachers can be shown how to use their professional judgments to look at the range of evidence about student learning and make a "holistic, integrative interpretation of collected performances." (Moss, 1994, p. 7) Reliability, then, becomes a judgment based on sufficiency of information rather than test-retest consistency.

Teachers can also be taught to develop public performance criteria that all students must apply to their work, even if they make their own choices about what work to do (see Figure 1 for an example). This level of standardization can allow for individual choice in projects and other types of performances while still ensuring that students' work will demonstrate their learning related to the targets of instruction. This will also help with rater consistency, the second dimension of reliability.

**Figure 1**

**Directions and Criteria for Literature Project**

This project will give you a chance to do some literary analysis. You will be working as a literary critic. In doing so, you will show your understanding of:

- how authors communicate major themes through their writing
- how authors communicate authors' their perspective or purpose in their writing
- how authors use language to create images, mood, and feelings
- how to judge the effectiveness of an author's work

You may choose a short story or a collection of three or more poems by a single author. In your writing be sure to include:

- a main message or theme you see in the story or poems
- what you believe is the author's purpose or perspective
- a description of at least two figurative language strategies the author used to communicate mood, images, and/or feelings
- specific examples from the story or poems to support your claims about theme, purpose, perspective, and figurative language
- an overall judgment about whether the author was effective in communicating themes and his/her perspective/purpose and in using figurative language strategies
- at least three reasons to support your overall judgment

If you choose to choose to use poems, make certain that the poems share a similar theme or message. Remember that authors often have more than one theme or message in their work, but be sure to focus your thinking on only one. Begin your paper by introducing the story or poems and the author. Organize your writing so that it will build a case for your positions and ideas about the writing. Look back at the literary reviews we have studied in class to give you ideas about how to organize your writing.

You must tell me what story or poems you have chosen to write about on _____________. You will turn in an outline or web for your paper on ______________. The first draft is due on ______________. Your final draft, the outline/web, and marked first draft are due on ______________. Be sure to give the source of the literary work(s) at the beginning of the paper.

*Reliability Dimension 2: Determining the degree of consistency in making decisions across students and across similar types of work.* Teachers generally use three types of assessment that could be affected by the consistency of their judgments about students' learning. They create short answer and essay items for tests; they assign projects and performances; they give several similar assignments (such as writing prompts) for which they have the same expectations. In these three situations, consistency of teachers' judgments depends on (a) whether the rules for scoring short answer items and essays are consistently applied across students, (b) whether the rules for scoring extended performances are applied consistently across students, and (c) whether rules for scoring frequently occurring types of assessment are applied consistently across similar tasks.

Teachers can be taught to develop public scoring criteria that they then apply to all students' performances. This can assist them in making consistent judgments across different students' performances. Teachers can be taught how to create generic scoring rules that can apply to multiple similar short answer or essay items (see Figure 2) so that they assess a range of responses to short answer or essay items based on the same criteria.

## Figure 2

### Generic Scoring Rules for Historical Essay

Performance Criteria

- Essay is clearly and logically organized.
- Position is clearly stated near the beginning of the essay.
- At least three arguments are given for the position.
- Arguments clearly support position.
- Specific supporting evidence is given for each argument.
- All supporting evidence is accurate and supports arguments.

Scoring Rubric

- **4 points** The essay is clear and logical in taking a position on a historical issue and in supporting the position with arguments and evidence. The essay thoroughly and effectively presents the position, arguments, and supporting evidence such that the reader can understand and entertain the writer's point of view. All supporting evidence is accurate.
- **3 points** The essay is clear and logical in taking a position on a historical issue and in supporting the position with arguments and evidence, although more evidence is needed for at least one argument. The essay presents the position, arguments, and evidence such that the reader can understand the writer's point of view. All supporting evidence is accurate.
- **2 points** The essay takes a position on a historical issue and supports the position with arguments and evidence, although more and/or stronger arguments and evidence are needed. The essay could be organized more effectively to communicate the position , arguments, and evidence. Some information presented may be inaccurate
- **1 point** The essay takes a position on a historical issue but provides little or no support for the position. Organization may or may not communicate the writer's ideas clearly. Some information presented may be inaccurate

If teachers learn how to frame the items and tasks given to students in a way that allows them to make consistent assessments and if they use scoring rules consistently across students and similar tasks, they are more likely to ensure that their evaluations of student's responses are consistent.

We have claimed in this paper that the frameworks we have set forth can be used to design assessment courses for teachers that not only better prepare them for the assessment tasks they will fact, but that help teachers develop habits of mind in which valid and reliable assessment is seen as central to the teaching-learning process. To support this claim, we briefly describe a course based on the validity and reliability frameworks and present evidence of its effectiveness.

## Assessment Frameworks in Action

The assessment course described here was taught at a large northwestern university, that provided a certification program for approximately 250 elementary and secondary teachers per year. Courses were ten weeks in length and a given class included pre-service teachers from all academic subjects and the arts for kindergarten through twelfth grade. During the quarter in which the assessment course was taught, students spent at least 20 hours per week in their field placement sites in addition to their course work as a transition into full time student teaching the following quarter.

During the summer of 1991, the decision was made to redesign the tests and measurement course for the teacher preparation program. Prior to that time, didactic procedures were used to cover standardized test interpretation, item writing and item analysis techniques, and statistical procedures for obtaining estimates of validity and reliability of tests. Students were assessed on their ability to write test items in various formats, and tested on their knowledge of measurement principles and concepts.

The redesign of the course was part of an overall restructuring of the teacher preparation program and was based on exit surveys indicating that students did not value the course (R. Olstadt, personal communication, May, 1991) as well as recommendations from the literature about what assessment courses for teachers should address (Airasian, 1991; Linn, 1990; Stiggins, 1991). In redesigning the course, the two most significant shifts were that (a) all assessment concepts were to be taught in the context of instructional practices and (b) the major emphasis of the course was to be on assessment validity and reliability rather than simply assessment techniques and memorization of abstract concepts.

We began with a model proposed by Linn (1990), and expanded it to include the use of process portfolios (Valencia, 1990; Wolf, 1991). We chose process portfolios because they are an interactive teaching tool in which successive iterations of work build upon one another to create a "prepared accomplishment" (Wolf, 1991), in this case a well developed plan that integrates instructional planning and assessment development using clearly defined learning objectives as the unifying force. We then planned assignments that would give students the opportunity to develop specific assessment literacy skills and strategies and that would require students to examine their own work in terms of validity.

In what follows we briefly discuss the work of the course and how the requirements of the assignments designed to help teachers develop the classroom-based definitions of validity and reliability given above. A more thorough description of the course is presented in Taylor and Nolen (1996) and Taylor (in press). In Taylor and Nolen (1996), each classroom course assignment is discussed in terms of its function in helping students think about one or more of the dimensions of validity, including excerpts from the students' self-evaluations that highlight the depth of their learning. In Taylor (in press), the types of decisions that had to be made to effectively use portfolios as an instructional and assessment tool are presented.

## The Process Portfolio

The portfolio provided both a means for instruction and learning during the course (process portfolio), and the product used to assess students' learning at the end of the course (showcase or assessment portfolio). The use of process portfolios allowed students to benefit from peer and teacher feedback (formative assessment) on the first draft of each assignment prior to its submission for grading purposes. Instructor feedback was intended to focus their thinking so that subsequent versions of their work reflected a better understanding of the course objectives. With better understanding, students could improve the quality of their own work.

At the end of the course, students pulled all of their work together in an assessment portfolio that "showcased" their learning for the course. They then wrote self- evaluations of their learning. In what follows the components of the of the portfolio are described.

## The Structure of the Assignments for the Course

To teach all five dimensions of validity and both dimensions of reliability, it was necessary to help students investigate assessment concepts in a meaningful context. The centerpiece of the course was a set of related assignments designed to guide students through the development of a unit of instruction so that they could engage in the thinking and skills necessary to make valid connections between learning objectives and instruction, instruction and assessment, and learning objectives and assessment.

For their assignments, students described a plan for a subject they would be likely to teach, and produced documents that were reasonable representations of the types of work good teachers do. Table 1 shows the assignments for the course and the dimensions of reliability and validity each was intended to help students learn.

### Table 1
### Configuration of the Portfolio Components for the Assessment Course

| Title | Description | Validity Dimension | Reliability Dimension |
|---|---|---|---|
| Subject Area Description | A description of the content foci and the instructional units in a subject area for an 8 to 12 week period including content coverage and major concepts targeted. | 1 | |
| Subject Area Goals and Objectives | 4-6 discipline based 4-6 objectives for each goal with discipline- based rationale for a subject the student planned to teach | 1 | |
| Instructional Unit Description | A description of instructional activities that would target 4-6 of the subject area objectives for 2-4 weeks of the period; with activities rationale indicating how each activity would help students learn the relevant objective(s) | 1, 3 | 1 |
| Item Sets:<br>• Checklist or Rating Scale<br>• Performance Assessment<br>• Essay Items<br>• Traditional Items | Four separate item sets as examples of the various types of assessment items and tasks that are used in classroom assessment (observational checklist or rating scale, performance assessment, essay items, traditional items (multiple choice, true-false, completion, matching, short-answer); each with the validity rationale | All | All |
| Sample Feedback | Mocked-up student work for one unit assessment with written or dialogue of oral feedback; philosophy and rationale about giving feedback | 5 | |
| Grading Policy | A description of the types of work that would be included in the grade, how different work would be evaluated, and how absences and late work would be handled; also included an example grade summary for one student | 1 | 1,2 |
| Self Evaluation | Description of own learning of selected course objectives, including discussion of concepts of validity, reliability, bias, and fairness referring to own work to show exam- ples of own learning | All | All |

Students were required to write rationales for all assessment decisions made during the development of components of the plan. Writing rationales forced students to articulate the validity and reliability issues that arose within each component of the plan, as well as giving the instructors a means to assess the conceptual learning that complemented the technical work displayed. The process of writing rationales also seemed to lead to deeper understanding of the concepts (Taylor and Nolen, 1996).

When all components were completed, students collected them into a final showcase portfolio. They wrote a single page reflection on each document and a self-evaluation of their learning in the course. In addition to these core assignments, other assignments were given to broaden students' understanding of assessment concepts. They included:

1. "Thought papers" in which they discussed their thoughts about collections of course readings (from the text book and a course reader).
2. A letter to parents explaining norm-referenced testing and score types
3. A written interpretation of one student's scores from a norm-referenced test.

The assignments listed above formed the core of the course as it evolved over the next twelve quarters. Based on student work and feedback, we adjusted the portfolio components, norm-referenced test interpretation assignments, and the number of thought papers required. We clarified instructions and experimented with various scoring schemes for the final portfolios. The focus of this paper is on the classroom assessment components of the portfolio; therefore, the latter three assignments are not discussed further here.

In what follows, we briefly discuss each of the components of the portfolio in the order the components were assigned. We also discuss the links between components and their links to the validity and reliability frameworks.

*Subject area description, goals and objectives.* Students began by writing a brief (one page) description of a subject area they planned to teach the quarter following the assessment course. The description included a general outline for one quarter or trimester, including the units of study and the major concepts and processes to be taught. The purpose of this component of the plan was to help students envision a subject area as a whole rather than as piece-meal units or text-book chapters. From this vision of the subject area, they were more able to articulate the overall learning goals of the course.

Once the general description was completed, students wrote four to six learning goals and four to six relevant objectives for those goals. We hoped that this level of objective writing would lead our students to clarify, for themselves, the most central learnings in the disciplines they planned to teach. This conceptual clarity is necessary if teachers are to develop assessments that reflect the disciplines studied (Validity Dimension 1).

Finally, students wrote a rationale describing how their goals and objectives reflected the substantive and syntactic structures of the discipline they intended to teach. This requirement built upon the educational psychology course they had taken the previous quarter in which they explored the concepts of disciplinary structure (Schwab, 1978) and pedagogical content knowledge (Grossman, Wilson, & Shulman, 1989). Students revisited this component throughout the quarter as they developed a deeper understanding of their goals and objectives through the assessment development process.

*Unit description.*Once students had completed their subject area descriptions, they described a brief unit of study that would fit within the quarter or trimester they had described in the subject area description. This component proved vital to students' understanding of how to establish the validity of assessments. Without the instructional unit as an anchor, it would be difficult to address aspects like the validity of methods of assessment for the methods of teaching used (Validity Dimension 3). Students developed units that were unique to their individual interests and that they were likely to use; therefore, the units were also a "hook" that kept students engaged in the work of the course..

Students selected up to six subject area objectives as the focus for the instructional unit. Then they wrote a brief narrative of the activities they would use to teach the objectives each day of the unit, linking the objectives to each activity, and providing a rationale for why the given activity or activities would lead to the targeted learning. This helped them to judge the fit of the assessments to the discipline as well as the fit of assessments to the unit of instruction (Validity Dimensions 1 and 3)

*Unit Assessments.* For the next part of the portfolio, students used a variety of techniques to create assessments for their instructional units. Students fully developed four different types of assessment for their units:

1. *An observational checklist or rating scale.* The assignment for the observational checklist or rating scale required students to identify one or more unit objectives and one or more situations from the unit for which observation would be an appropriate form of assessment. The checklist or rating scale was to have at least 10 items that were of clearly observable behaviors that could show the learning described in the objective(s).
2. *A performance-based assessment.* This assignment included a description of a performance that was either an integral part of the instructional unit or that could be used for students to show the learning objectives that were the target of the instructional unit. Students wrote directions (oral or written) that were sufficient for their students to complete the performance and show the learning, as well as a checklist, rating scale, or rubric(s) to evaluate the performance.
3. *Two essay items.* The assignment for the essay items required students to think about two essay prompts that could be written in the instructional unit through which students could show learning related to one or more of the unit objectives. Essay prompts had to be explicit enough that students would know what they were to do to successfully write the essays. Essays were to be brief (extended essays were considered performance assessments). Students also had to write scoring rules (checklists, rating scales, and/or rubrics) for each essay.
4. *A set of "traditional" test items.* This assignment was for a set of at least 10 items that assessed one or more unit objectives. The set had to include at least three multiple-choice items, one matching item, two completion item, two true-false items, and two short answer items. The item set could be organized as a quiz, part of a unit test, or into one or more daily worksheets (for younger students). Students had to develop a scoring key for the select items and scoring rules (key words, checklists, rating scales, or rubrics) for the supply items.

Students were asked to develop assessments that fit with their instructional methods and that assessed their unit objectives. Students then had to write a rationale for each item or task that answered several questions:

1. How will the item/task elicit students' learning related to the targeted unit objective(s)? (Validity Dimensions 1 and 2)
2. How does the item/task reflect concepts, skills, processes that are essential to the discipline? (Validity Dimensions 1 and 5)
3. How does the item/task fit with the instructional methods used in the unit? (Validity Dimension 3)
4. How do the rules for scoring the item/task relate to the targeted unit objective(s)? (Validity Dimension 1)
5. Is the mode of assessment such that all students who understand the concepts will be able to demonstrate them through the assessment? (Validity Dimension 4)

By thinking about each item or task and its relationship to the discipline and the unit methods, students went beyond simply practicing item or task writing techniques and had to consider whether the assessment represented the construct (Validity Dimension 1) and whether the assessment was appropriate for the instructional context (Validity Dimension 3). By examining whether items and tasks clearly asked for the learning targeted, students could examine whether assessments were presented in a way that allowed their students to demonstrate learning (Validity Dimension 2; Reliability Dimension 1). By carefully examining the rules for scoring the item/task and how these rules relate to the objective(s) the item/task is intended to measure, students had to think about whether the scores used to represent student performance related to the construct (Validity Dimension 1) and whether their scoring rules would help them be more consistent across students (Reliability Dimension 2). By having to discuss whether all of their students would be able to show their learning through the mode of assessment, our students could begin to explore issues related to bias (Validity Dimension 4). By considering the link between the assessments and the disciplines, students could also begin to grapple with whether assessments were likely to provide appropriate representations of the disciplines for students (Validity Dimension 5). Finally, by creating several assessments in different modes for the same unit and unit objectives, they were able to compare different methods of assessment in terms of their demands for students (Validity Dimension 2).

*Feedback.* This assignment required students to choose one of their assessments and either try it out with one of their students or mock-up/describe one of their students' responses. They then showed what they would do (either by marking on the paper or by describing a dialogue with their student) to give feedback. Finally, they wrote a rationale for the feedback, including both a discussion about the influence of the feedback on the learner's motivation and self-esteem and a discussion about how the feedback could help their student improve future performance related to the learning target(s). This gave students another opportunity to explore the consequences of assessment interpretation and use (Validity Dimension 5).

*Grading Policy.* For the grading policy, we had students use the assessment ideas derived from their unit plans and write a grading policy that applied to the entire subject area description. They had to choose an grading philosophy (norm-referenced or criterion-referenced) and explain why they had chosen it. They explained what types of work would contribute to the grades (e.g., essays, reports, projects, tests, homework, daily seatwork, etc.) and why this work was important to learning the discipline (Validity Dimension 1), the general strategies they would use to assess various kinds of work (Reliability Dimension 2 [e.g., a generic four point rubric for all homework assignments based on completeness and accuracy of work]), how they would weight the various sources of assessment information, and how they would summarize across assessments to assign a grade. They also had to prepare a sample grade summary for one student using the information from the various assessment sources.

Students had decide how much weight to give to attendance, timeliness, oral participation, and attitude when making judgments about their students' learning of the targeted objectives. By validity standards, some of these variables would be considered sources of irrelevant variance that lead to invalid inferences about student academic learning (Validity Dimension 1). They also had to think about multiple-sources of evidence needed to make reliable decisions about learners (Reliability Dimension 1). Finally, by creating a set of scores for a hypothetical examinee, they were able to look at the impact of various sources of assessment information on overall grades (Validity Dimension 5)

*Reflection and Self-evaluation.* The final component of the portfolio was the self-evaluation. This component gave students an opportunity to bring closure to the course and to organize their thinking about a few central assessment concepts using the work required in the course as the anchor. In these self-evaluations, they wrote about their understanding of major assessment concepts for the course. They were required to:

1. Discuss their current understanding of the concepts of validity, reliability, bias, and fairness with reference to specific work in the course that helped them understand these concepts and *how* the course work had helped them to understand the concept.
2. Select at least six of assessment course objectives and discuss what they had learned related to each objective, what aspect of the course had helped them to learn it, and how.

The self-evaluations were evaluated for the students' ability to demonstrate their understanding of the assessment concepts using their work as examples. It was not sufficient to provide a text book definition of a term or to explain the impact of assessment in general terms; specific and credible examples were required. In the following discussion, excerpts from student self-evaluations from the Spring 1994 students are used to demonstrate, in their own words, what students thought about as they reflected on their own learning. Selected excerpts represent common thoughts among students.

In the self-evaluations, when students discussed their understandings of validity, most references were made to the unit assessments (Validity Dimensions 1 through 5). Discussions of reliability and fairness usually focused on the use of rubrics and observational checklists and rating scales (Reliability Dimension 1 and Validity Dimension 4). Rarely did students bring up consistency of ratings across students and performances as an element of reliability (Reliability Dimension 2). In discussions of fairness and bias, students often indicated how helpful it had been to use a standardized scoring scheme to evaluate essays or performances in class; how such rules had given them a way to be fair and unbiased in their assessments (Validity Dimension 4). For example:

> "The students in my placement are intentionally given vague criteria. The teacher considers it her right to use her personal judgments of the student's attitude and behavior to influence the grade. If the criteria (are) not spelled out she has the leeway to insert her prejudice. Students realize what is going on and they become cynical and resigned. Few of them try to fight it. This lack of fairness is so widespread that they have come to expect it."

When choosing which component of the portfolio most influenced their learning, each component was selected by someone. For some, the clarification of their disciplines were seen as the most critical element (Validity Dimension 1).

> "The best part of the course for me was the subject area description and goals because it forced me to stop and think about why I want to teach biology. . . . Being a good teacher is a difficult task. The best way to overcome this is going through the process we went through during the development of subject description, goals, objectives, and rationale. . . . It will help me down the road as a teacher."

Some students wrote about the importance of developing a unit of instruction in order to help them conceptualize the role of assessment (Validity Dimension 3).

> "It made me focus on what I really wanted my students to learn, and then I had to find different and appropriate ways to assess whether or not the students learned these things. If one of my unit objectives was to view the American Revolution and its effects from a variety of perspectives, then an assessment that only deals with one perspective is not a valid assessment. It does not tell me if they have learned what . . . I want them to learn."

Many students chose to focus on one or more of the unit assessments, discussing what they had discovered as they developed a given type. A very common observation was about the need for clear directions for performances so that their students would actually provide performances that showed the targeted learning (Validity Dimension 2).

> "Giving the criteria for successful work helps make an assessment valid, as it insures that a student's essay demonstrates the student's conceptual and/or procedural understanding rather than his/her ability to read the teacher's mind."

Another common focus was on the fit between various forms of assessment and either the discipline or the learning objectives as well as what assessments communicate to students about a discipline (Validity Dimensions 1 and 5).

> "Assessments are not neutral!. . . Assessments send messages about a discipline; they communicate to students in a direct, concrete, and powerful way about what is really important to know is this subject."

Students also wrote about grading policies. They typically reflected back on readings about the influences of grading practices on motivation and self-esteem (Covington & Beery, 1976; Canady & Hotchkiss, 1989), discussing the assumptions often made about the motivating power of grades and considering the potential consequences of various ethical and unethical grading practices (Validity Dimension 5). Some students indicated that in being forced to think about the relative weight of each aspect of the grade, they had to look again at the discipline to decide which sources of evidence were best and most important in making judgments about their students' learning (Validity Dimension 1).

These and other comments showed us, as instructors, the power of the work assigned in the course in terms of helping our students understand important assessment concepts. Comments from students suggested that the assignments done for the course as well as the rationales and self-evaluation enhanced their learning.

## Comparative Studies of the Traditional Tests and Measurement Assessment Course and the Portfolio-Based Course

In an effort to evaluate the effectiveness of the revised course, three studies were conducted that compared data available from students who had taken one of the two versions of the course: the portfolio-based course and the traditional tests and measurement course. The classroom assessment component of the original assessment course covered item writing and item analysis techniques (some later sections of this course also covered performance assessment), and statistical procedures for obtaining estimates of validity and reliability of tests. Instructors used a combination of lectures and discussions to teach assessment content. Instructors relied heavily on midterm and final examinations (primarily multiple-choice), which counted for 60 to 70 percent of the final grade (depending on the instructor). Up to 25 percent of the final grade was based on students' development of behavioral objectives (based on Bloom's taxonomy) and tests or sets of items to measure those objectives. Tests or sets of items were independent of any context except that of the behavioral objectives.

Study 1 compared course evaluations across teaching faculty for the two versions of the course. Study 2 compared evaluations of relevant components of an exit survey given to all students graduating from the teacher education program. Study 3 involved analyses of data from follow-up surveys sent to teacher education students in the quarter following their enrollment in the assessment course--the time during which most were engaged in full-time student teaching. In the survey, the pre-service teachers were asked to discuss assessment issues, validity dilemmas, and reliability dilemmas that arose in their teaching. Each of these studies is described more fully below.

The designs of the three studies reflect the natural development of curricular revision, rather than the carefully-controlled world of laboratory studies or field experiments. The research opportunity was presented by the decision to redesign the course. Thus, comparisons of the two versions of the course presented in Studies 1 and 2 depended on existing institutional data. The data for Study 3 were collected as part of an evaluation of the course revision, but the decision of one instructor to revert to the traditional format for two sections provided an opportunity for comparison on the follow-up measure.

### Study 1: Course Evaluations

*Data Source*. The university's Office of Educational Assessment provided course evaluation results for each quarter from the summer quarter of 1988 through the spring quarter of 1994. Course evaluations are required for every course for assistant professors and at least once a year for senior faculty. Student participation is voluntary, however, most students complete the form. Results of the course evaluation are not given to the instructor until after grades are submitted.

Data representing 12 quarters of the traditional tests and measurement version of the course and 12 quarters of the revised course were available. The number of respondents from the traditional tests and measurement course ranged from 15 to 55 across different sections with a mean of 32.25. The number of respondents from the revised course ranged from 17 to 74 with a mean of 32.58. Because responses were anonymous, it was not possible to determine the exact number of males and females in the sections nor the number of students who were to be certified in elementary, secondary, or music education. Academic ranks for the instructors in the traditional tests and measurement course ranged from graduate student instructor to full professor. Academic ranks for the instructors in the revised course ranged from graduate student instructor to assistant professor. There were 8 different instructors for the traditional tests and measurement course and 3 different instructors for the revised course.

Only those items common to evaluation forms used in all sections of the course were included in the analysis. Items common to all forms are given in Appendix A. Each item was rated on a 6 level scale. "Excellent" (5), "very good" (4), "good" (3), "fair" (2), "poor" (1), and "very poor" (0). Four items from this common set assessed students' ratings of the content and relevance of the course.

*Results*. Mean item scores were averaged across classes for each version of the course. Only those items specifically related to the content of the course and the relevance of the course were included in the analyses. Two analyses were performed on a selected set of the items. In the first analysis, data from four items from the course evaluation forms were used: (a) course as a whole, (b) course content, (c) amount you learned in the course, and (d) relevance and usefulness of course content.. These items were summed to obtain an overall score for the general content of the course; the mean score for the traditional tests and measurement course was 12.09 (SD = 2.04), and for the revised course was 16.48 (SD = 1.62). In the second analysis, relevance and usefulness was analyzed alone, with means for the traditional tests and measurement and revised course 2.92 (SD = .57) and 4.29 (SD = .38), respectively.

T-tests were performed to compare mean ratings for these data. There were significant differences between students perceptions of the general content of the course (t(22) = 5.85, p < .001) and the relevance and usefulness of the course (t(22) = 7.00, p < .001). Students in the revised course saw the course as more relevant to their needs and rated the content of the course between "very good" and "excellent." Students in the traditional tests and measurement course rated the course as "good" on both general content and relevance and usefulness.

These differences might have been due to differences in the effectiveness of individual instructors. However, even instructors of the traditional tests and measurement course who received high ratings for instructor's effectiveness received lower ratings on relevance and usefulness of course content. and course content. Two instructors from the traditional tests and measurement course had high ratings for instructor's effectiveness (mean ratings of 4.38 and 4.25), comparable to the average ratings for the two revised course instructors with the highest effectiveness ratings (mean ratings of 4.20 and 4.54). When only these four instructors are compared, the mean ratings for the for relevance and usefulness were 3.52 and 3.83 for the traditional tests and measurement course and 4.81 and 4.71 for the revised course. The mean ratings for course content were 3.90 and 3.64 for the traditional tests and measurement course and 4.71 and 4.54 for the revised course. This suggests that whether students saw the content of the assessment course as relevant to their needs was not merely a function of their perceptions of the effectiveness of an instructor.

## Study 2: Teacher Education Program Exit Surveys

*Subjects*. As part of the ongoing evaluation process of the teacher education program, exit surveys were administered in the last quarter of the program to all students. We obtained 153 of these surveys from three years just prior to the change in the assessment course (1989-91) and 145 from two years after the change (1992 and 1994). In the summer of 1992 an outside instructor taught a traditional tests and measurement course. Since it was not possible to tell which 1993 exit surveys came from students who had taken the revised course, data from that year were not used. All responses were anonymous; therefore, the demographic characteristics of the respondents were unavailable.

*Instruments*. Exit surveys were general program review instruments and asked a variety of questions about students' experiences in the teacher education program, including both course work and field work. There were several items which provided some information about students' perceptions of assessment course effectiveness. First, a set of items asked students to rate how well the program as a whole had prepared them in a number of areas corresponding to the state requirements for teacher education programs. One of these items was "How well has this program prepared you to evaluate student work," which students rated on a scale from 1 ("not at all prepared") to 5 ("thoroughly prepared").

A set of open-ended questions asked students to comment on various program aspects. Three of these questions were coded for comments related to the assessment course.

The first of the open-ended questions asked for comments about any of the courses in the program. Coding schemes for this item were as follows:

1. Comments specifically directed at the assessment course, and related to value or worth of the course or its content were coded (0) if they suggested eliminating the course altogether; (1) if they stated the course was worthless, not valuable, not useful for teachers; and (2) if they stated the course was valuable, applicable or useful.
2. General comments (not referring to value) were coded (1) negative or (2) positive.

A second item asked students to list aspects of the teacher education program that were particularly valuable or worthwhile. Raters counted the number of students listing the assessment course here.

A third item asked what important material was left out or not sufficiently covered. Raters counted any mention of an assessment-related topic (e.g., setting up grade books, portfolios, informal observation). Finally, negative comments regarding the work load related to the assessment course mentioned anywhere in the survey were counted.

All coding was completed by the authors and one graduate student who was unfamiliar with the purpose of the research. There was a 98% agreement among the three raters. Final counts for each code assigned to each response were based on absolute agreement among the raters.

*Results*. Ratings of how well students thought the program prepared them to do assessment were compared across courses using a one-way ANOVA. Students who took the revised course rated the teacher education program as preparing them more thoroughly to do assessment (Mean = 4.07, SD = 0.87) than did students who took the traditional tests and measurement course (Mean = 3.22, SD = 1.04; $F(1, 296)= 58.36$, $p < .001$).

Frequency of responses for each open-ended item appear in Table 2. In general, the comments were more positive for the revised course, though not uniformly so. Typical comments for the traditional tests and measurement course included "[The assessment course] was a useless class. Testing and evaluation are essential, but I learned almost nothing in this class" and "Did not relate to the real world." Typical comments for the revised course included "[The assessment course] provided me with the information that I considered most valuable in my field experience" and "[The assessment course] was the most valuable class overall for my teaching." Eight students in the revised course (5.2%) stated that the work load in the revised course was excessive, while none of the students taking the traditional tests and measurement course did so.

### Table 2
### Frequency of responses to each item for the traditional tests and measurement course and the revised course

**Comments (Value)**

| Course | N of Cases | Valuable | Not Valuable | Eliminate Course |
|---|---|---|---|---|
| **Revised Course** | 145 | 19 | 2 | 0 |
| **Traditional Course** | 153 | 1 | 17 | 9 |

**Comments (General)**

| Course | N of Cases | Positive | Negative | Negative Work load |
|---|---|---|---|---|
| **Revised Course** | 145 | 22 | 4 | 8 |
| **Traditional Course** | 153 | 0 | 9 | 0 |

**What aspects of the program were...**

| Course | N of Cases | Particularly Valuable | Not Sufficiently Covered |
|---|---|---|---|
| **Revised Course** | 154 | 28 | 3 |
| **Traditional Course** | 129 | 1 | 11 |

Each comment was coded into only one category, but some students mentioned the assessment course in more than one way. Therefore a new variable was created by counting the number of students in each group who had responded in some way that the assessment course was valuable and the number of students who had indicated that the course was not valuable. Students who had taken the revised course were much more likely to mention it as a valuable part of the program (31%) than to say it was not (2%), while those taking the traditional tests and measurement course were more likely to see the course as not valuable (17%) than as valuable (1%) (chi-square(1) = 61.8, $p < .001$).

## Study 3: Follow-up Survey

Study 3 aimed to assess the impact of the assessment courses on pre-service teachers' work in their field placement classrooms. We were primarily concerned with their ability to describe assessment issues they faced in teaching, and in their understanding of validity and reliability. We were also interested in the extent to which they could use the assessments (and other components of their work for the course) in their field placement classrooms.

*Subjects*. Students from six different quarters were asked to be part of an anonymous mail survey during quarter following the one in which they took the assessment course. Most of the students were engaged in full-time student teaching. Two classes of students (N = 112) who had taken the traditional tests and measurement course during the summer of 1992 were surveyed. Twenty-one percent (n = 23) of these students completed and returned the surveys. Five classes of students (N = 195) who had taken the revised version of the course between the summer of 1991 and the autumn of 1992 were surveyed. Twenty-five percent (n = 50) of those enrolled completed and returned the surveys.

*Results*. The follow-up questionnaire addressed a number of assessment and programmatic issues. A complete list of items is shown in Appendix B. There were few differences between groups on the assessment methods used in their field placement, the proportion of planning time spent on assessment, or the amount of time they reported thinking about assessment. Students in the revised course reported spending slightly more time planning assessments (7% of planning time, SD = 4%) than traditional tests and measurement course students (3%, SD = 4%), t(65) = 9.54, p <.01).

Ninety-two percent of the students in the revised course reported using all or part of the work developed in the course, while only 8% of students in the traditional tests and measurement course reported using any of the work developed in their course (chi-square(1)=9.03, p < .01). Students who reported using materials developed in the course rated the process of planning helpful on a 5-point scale from 1 ("not at all helpful" to 5 "very helpful"), with a mean of 4.17 (SD = .81).

Three items provided information on students' post-course understanding of assessment issues, validity, and reliability. Responses to items 4, 6, and 7 (the influence of assessment, validity issues, and reliability issues) were independently coded by three full professors with strong measurement and statistics backgrounds who had previously taught classroom assessment courses. They were not aware of the purposes of the study or the type of course in which students were enrolled. Coding was based on the degree to which the students' responses showed understanding of general assessment concepts. Table 3 provides the scheme used to code student responses.

## Table 3

**Coding scheme for relevant items of the post-course survey**

4 Influence of course on teaching
Code 1: 1 = yes 2 = no
Code 2

- 2 = shows clear, unambiguous understanding of appropriate uses of assessment
- 1 = - shows partial understanding of appropriate uses of assessment
  - describes delivery of instruction; may have assessment links
  - uses assessment terms without examples
- 0 = shows little or no understanding of appropriate uses of assessment in instruction
- NS = not scorable (off task or omitted)

6 Validity issues

- 2 = gives good example of validity issue
- 1 = - possible example of validity issue, somewhat unclear
  - may confuse validity with reliability
- 0 = gives example that is neither reliability nor validity
- NS = not scorable (off task or omitted)

7 Reliability issues

- 2 = gives good example of reliability issue
- 1 = - possible example of reliability issue, somewhat unclear
  - may confuse validity with reliability
- 0 = gives example that is neither reliability nor validity
- NS = not scorable (off task or omitted)

The final code assigned to each item for each examinee was based on a majority agreement among the raters. For students from the traditional tests and measurement group, 35% indicated that the course had no effect on their teaching. For the students in the revised course, 2% indicated that the course had no effect on their teaching.

For influence of assessment course, 70 percent of students from the revised course showed a clear understanding of the appropriate uses of assessment, as judged by the raters, as compared to 44 percent of the students in the traditional tests and measurement course (chi-square(3) = 9.96, $p < .02$). For validity issues, 70 percent of students from the revised course gave good examples of validity issues as compared to 22 percent of the students from the traditional tests and measurement course (chi-square(3) = 15.01, $p < .001$). For reliability issues, 22 percent of students from the revised course gave good examples of reliability issues as compared to 13 percent of the students from the traditional tests and measurement course (chi-square(3) = 8.74, $p < .03$), however, a fairly large proportion of both groups gave no examples at all (61% of the traditional tests and measurement course students and 42% of the revised course students). In addition, a fairly large percent of the students in the revised course (32%) received a score of 1 for this item, indicating that while the students in the revised course may have been better prepared to address issues related to reliability than were the students in the traditional tests and measurement version of the course, they were still not sufficiently prepared.

## Discussion

The results of these three studies suggest that the revision of the assessment course was beneficial to preservice teachers. Students taking the revised course were more likely to see the course as useful and relevant to their own work as teachers than students in the traditional tests and measurement course, both at the end of the quarter in which they took the course, and at the end of their teacher education program (following full-time student teaching). Students taking the revised course felt better prepared to deal with classroom assessment than similar students in the traditional tests and measurement course by nearly a full standard deviation. Nearly a third of those responding listed the revised course as one of the most useful parts of the teacher education program; only 1% of students listed the traditional tests and measurement course.

Although student ratings are valuable, they do not bear on whether students actually learned central concepts in assessment and could use those concepts in their own classrooms. The results of the follow-up survey (Study 3) suggest that students in the revised course were indeed able to use the notion of validity generatively. The concept of reliability; however, was not as clearly understood by the majority of students in either version of the course.

Post-course questionnaires showed that, while students in the revised version of the course had a better understanding of reliability (as it applied to the assessments used in their field placements) than did students in the traditional tests and measurement course, their understanding of reliability was still inadequate. This could be due to the intense focus on a broad understanding of validity and inadequate attention to reliability issues. Many of the examples of reliability issues given by students who had taken the revised version of the course were actually validity issues. These comments, while inaccurate representations of the concept of reliability, did show an understanding of the difference between appropriate and inappropriate assessment practices. On the other hand, survey comments from students who had taken the traditional tests and measurement class indicated that they were very confused about meanings of reliability and validity. Several of these students responded to the questionnaire items about reliability and validity with:

> " I don't understand the concept. I only memorized it for class."

It appears from these data that the revised assessment course was effective in helping students understand appropriate assessment practices in the context of the classroom and in helping them develop a generalizable understanding of the concept of validity. What was lacking was a deep understanding of reliability and how it transferred to the world of teaching. Subsequent to these analyses, the course was revised in order to help students focus more carefully on the dimensions of reliability. Follow-up studies are planned to determine whether these adjustments accomplished the course goals.

## Conclusion

The assessment course outlined here has been designed to engage students in tasks relevant to their own work as preservice teachers and demand that they consider assessment in the context of disciplinary structures and instructional practices. Each component of the portfolio gave students an opportunity to address one or more of the dimensions of validity and reliability highlighted in this paper. The focus on validity guided student learning from the initial subject area description and concomitant goals and objectives (which helped students develop clearer definitions of their disciplines for themselves), to the unit assessments (which helped students explore all five dimensions of validity), to the grading policy (which helped them address issues of multiple sources of evidence, appropriateness of evidence, and potential consequences of assessment interpretations and use).

One powerful aspect of this course may have been that it was a model of the concepts students were learning. In contrast to a course in which teachers act as impartial observers of students' learning, the instructors were engaged as participant observers--using feedback and guidance to help ensure learning for as many students as possible. Multiple sources of information were used to determine whether students were learning the concepts and skills of the course, from the components of the portfolio to the reflections and self-evaluations at the end of the course.. Students had more than one opportunity to return to their work and revise based on feedback from the instructor and later learning. As such, the instructors had multiple opportunities to observe students' growth over time. Public criteria were used to communicate the expectations of performances and scoring rules were consistently applied across students' work and across similar performances.

Another powerful aspect of the course was that it was carefully focused on tasks and reflective writing designed to help students grapple with each of the dimensions of reliability and validity described in this paper. The learning that resulted from the course--in terms of students' transfer of ideas from the course to their own teaching as judged by three full professors with substantial knowledge of assessment concepts--suggests that the validity framework used to organize the work of the students is one that teachers can internalize and understand. A stronger focus on the sufficiency of assessment information and ways of ensuring scoring consistency in students' work was needed if students were to better understand the concept of reliability.

The success of this course in reaching teachers has implications not only for the preparation of teachers, but for the ways in which we present measurement theory in textbooks and instruction and for how classroom assessments are used in large scale assessment programs. While there may be a place for external assessments that provide accountability data to taxpayers, legislators, and state boards of education, the measurement model developed for these external tests does not fit the rich and complex environment in which learning takes place.

If we are to adequately prepare teachers in the area of assessment, clearer thinking is needed about the assessment concepts, types of textbooks and the methods of teaching that are used. Measurement professionals often lament the wide-spread lack of understanding about measurement concepts. Quite possibly we have created this problem ourselves. The problems seen may be due to the fact that the philosophical foundations of test theory don't fit the classroom context well. Although text book authors may be trying, in their individual ways, to construct texts book that will force a fit where one does not exist, we may need to admit that a test theory that fits the modernist notions of the impartial observer is not appropriate for the context in which the teaching and learning occur.

It is likely that two different frames are needed for educational assessment constructs: one for the context of school and one for the context of external norm-referenced tests. Textbooks could acknowledge the differences between these contexts and frame concepts, procedures, and skills as appropriate for each context. Courses could be designed to help teachers internalize and grapple with these differences. Textbooks and teacher educators could regularly bring teachers back to classroom-relevant dimensions of validity and reliability within chapters that address various assessment problems, skills, decision-making issues and processes for the classroom. They could ask students to think deeply about why very different frameworks and methodologies apply to external assessments. As measurement professionals and teacher educators, we could do a better job of preparing good "participant observers," as well as helping teachers understand the paradigm shifts between the two perspectives on assessment. Most importantly, we should frame our preparation of teachers in such a way that they are clear about their own tasks as teachers: to promote students' ongoing learning.

At this point in time, while we have standards for educational and psychological testing (AERA, APA, NCME, 1985), standards for assessment competencies for teachers (AFT, NCME, NEA, 1990), and standards for various professional groups in the interpretation and use of tests (e.g., American Association for Counseling and Development, 1989; APA Committee on Children, Youth and Families, Committee on Testing and Assessment and Committee on Ethnic Minority Affairs, 1992; American Speech-Language-Hearing Association, 1991), we do not have standards for the preparation of teachers related to assessment or for the materials used in that preparation. In addition, as AERA, APA, and NCME revise the testing standards, it is critical that they look carefully at the contexts in which assessments apply as well as the philosophies underlying the use of assessments within those contexts rather than attempting to create omnibus standards that apply to all assessment circumstances.

Related to this, as large scale assessment programs look at the viability of incorporating classroom-based assessments into statewide accountability information, the nature of the classroom context, and the proposed validity and reliability frameworks, should be considered. Some might say that, given the unstandardized and progress-oriented nature of classroom assessments, the information derived from these sources is too unstable to use for large scale assessment purposes. On the other hand, the richness and breadth of the assessment information that arises from classrooms could give us more and better information if we more appropriately develop teachers' assessment skills.

As state and national programs attempt to incorporate classroom assessment information when reporting on students' learning, the focus must be on the validity and reliability frameworks that fit the classroom rather than ones that fit external tests. The dimensions of validity and reliability presented here make sense to teachers because they make sense in a classroom context of teaching and learning. Large scale assessment programs that use classroom-based evidence should consider the dimensions of validity and reliability relevant to the classroom when making decisions about how to incorporate classroom-based information into large-scale programs.

If, in order to obtain assessment information from classrooms, large scale programs create top-down standardized tasks or tests to be administered by teachers, the validity of such assessments for the classroom context is suspect. Given the validity framework presented here, top-down classroom assessments could not provide valid classroom assessment information because they would not follow from instruction (Validity Dimension 3). They would simply be extensions of external, standardized tests. If teachers are admonished to use standardized administration directions that do not allow for the unique needs of students, top-down classroom assessments should be suspect because they may prevent some students from showing their learning in ways that accommodate their unique needs (Validity Dimension 4). If standardized, top-down tasks are closely circumscribed in order to strengthen reliability, they limit the capacity of the assessments to assess students' understandings of the subject area disciplines (Validity Dimension 1). This would not only limit fit with the content and constructs to be measured, but would rob the classroom of the opportunity to use important assessments to accurately represent the structures of disciplines (Validity Dimension 5). Limiting classroom assessment information to a few, standardized, top-down assessments would also limit the range of evidence and counter-evidence that teachers could present about student learning--a threat to Validity Dimension 2.

If, on the other hand, several generic outlines for tasks, scoring rules, and tests are created. (e.g., Rekase, 1995), and teachers are allowed to configure these assessment outlines to fit their own instructional methods, content focuses, and timelines, classroom assessments could fit all of the dimensions of validity relevant to the classroom context. Guidelines for adaptation of the assessments to instructional contexts, strategies for evaluating the validity of these adapted assessments, and ideas about what would constitute a reasonable range of assessment information for decision-making could help teachers develop useful assessments, first for themselves and their students and secondly for large scale programs. State programs could provide powerful professional development materials to practicing teachers through such materials.

For too long, rules for creating and evaluating external tests have been seen as the ideal for obtaining valid and reliable information about learning in the classroom. This has led to a lack of fit between the needs of teachers and the notions of assessment professionals. With the current awareness of the importance of assessment among teachers, school administrators, and policy-makers, the classroom has the potential to be a much more powerful and complete source of assessment information. To achieve this potential, however, we must begin with frameworks for measurement constructs that fit the classroom context, teach teachers how to use these frameworks to improve the quality of their assessments, and ensure that external uses of classroom assessment information attend to these frameworks when deciding how to incorporate classroom assessments into large scale programs.

## References

Airasian, P. (1991). Perspectives on measurement instruction for pre-service teachers.

Educational Measurement: Issues and Practice, 10 (1) 13-16, 26.

Airasian, P. (1991). Classroom Assessment, First Edition. New York: McGraw-Hill.

Airasian, P. (1993). Classroom Assessment, Second Edition. New York: McGraw-Hill.

American Association for Counseling and Development. (1989, May 11). The responsibilities of test users. Guidepost, 12, 16, 18, 27.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students, Washington, DC: Author.

American Psychological Association Committee on Children, Youth and Families, Committee on Testing and Assessment & Committee on Ethnic Minority Affairs. (1992). Psychological testing of language minority and culturally different children, Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing, Washington, DC: American Psychological Association.

American Speech-Language-Hearing Association. (1991). Code of ethics of the American Speech-Language-Hearing Association. Rockville, MD: Author.

Anderson, L., Blumenfeld, P., Pintrich, P. I., Clark, C., Marx, R., & Peterson, P. (1995). Educational psychology for teachers: Reforming our courses, rethinking our roles. Educational Psychologist, 30, 143-157.

Bloom, B. S., Madaus, G.F., & Hastings, J. T. (1981). Evaluation to Improve Learning. New York: McGraw-Hill Book Company.

Bricklin, B., & Bricklin, P. M. (1967). Bright child--poor grades. New York: Dell.

Butler, R., & Nisan, M. (1986). Effects of no feedback, task related comments, and grade on intrinsic instruction and performance. Journal of Educational Psychology, 78, 210- 216.

Canady, R. L., & Hotchkiss, P. R. (1989). It's a good score! Just a bad grade. Phi Delta Kappan, 71 (1), 68-71.

Cohen, D., & Ball, D. L. (1990). Relations between policy and practice: A commentary. Educational Evaluation and Policy Analysis, 12 (3), 331-338.

Covington, M. V., & Beery, R. G. (1976). Self-worth and school learning. New York: Holt, Rinehart, 42-63, 77-87.

Covington, M. V., & Omelich, C. L. (1984). Task-oriented versus competitive learning structures: Motivational and performance consequences. Journal of Educational Psychology, 76, 1199-1213.

Cronbach, L. J. (1970). Essentials of Psychological Testing, Third Edition. New York: Harper & Row, Publishers.

Deci, E., & Ryan, R. (1985). Intrinsic motivation and self- determination in human behavior. New York: Plenum.

Deci, E., & Ryan, R. (1987). The support of autonomy and the control of behavior. Journal of Personality and Social Psychology, 53, 1024-1037.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4 (4), 289-303.

Galton, F. (1889). Natural inheritance. New York: Macmillan.

Glesne, C. & Peshkin, A. (1992). Being there: Developing understanding through participant observation. Becoming Qualitative Researchers: An Introduction. White Plains, NY: Longman, 39-61.

Grossman, P. L. (1991). Overcoming the apprenticeship of observation in teacher education coursework. Teaching and Teacher Education, 7 (4), 345-357.

Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. C. Reynolds (Ed.), Knowledge base for the beginning teacher. New York: Pergamom.

Gullickson, A. R. (1986). Teacher education and teacher- perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23 (4), 347-354.

Gullickson, A. R. (1993). Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs, and challenges. In S. L. Wise & J. C. Conoley (Eds.), Teacher training in measurement and assessment skills. Lincoln, NE: Burros Institute of Mental Measurements, University of Nebraska.

Hanna, G. S. (1993). Better Teaching Through Better Measurement. Fort Worth, TX: Harcourt Brace Javanovich College Publishers.

Hieronymus, A. N., & Hoover, H. D. (1987). Iowa Tests of Basic Skills: Writing Supplement Teacher's Guide. Chicago: Riverside Publishing Company.

Houghton Mifflin Company, American Heritage Dictionary of the English Language, Morris, W. (Ed.). Boston: Author, 1098.

Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. Teachers College Record, 91 (3), 422-436.

Linn, R. L., & Burton, E. (1994). Performance-based assessment: Implications of task specificity. Educational Measurement: Issues and Practice, 13 (1), 5- 8, 15.

Linn, R. L., & Gronlund, N. E. (1995). Measurement and Assessment in Teaching, Seventh Edition. Englewood Cliffs, NJ: Merrill, an imprint of Prentice Hall.

Lortie, D. (1975). Schoolteacher: A sociological study. Chicago: University of Chicago Press.

Mehrens, W. A., & Lehmann, I. J. (1991). Measurement and Evaluation in Education and Psychology, Fourth Edition. Fort Worth, TX: Holt, Rinehart, and Winston, Inc.

Messick, S. (1989). Validity. In Educational Measurement. Robert Linn (Ed.). Washington, DC: American Council on Education.

Moss, P. A. (1994). Can there be validity without reliability? Educational Researcher, 23 (2), 5-12.

Moss, P. A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. Educational Researcher, 25 (1), 20-28.

National Council of Teachers of Mathematics (1991). Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions. Reston, VA: Author.

National Council of Teachers of Mathematics (1995). Assessment Standards for School Mathematics. Reston, VA: Author.

Nicholls, J. G. (1989). The competitive ethos and democratic education. Cambridge, MA: Harvard University Press.

Nitko, A. J. (1996). Educational Assessment of Students. Englewood Cliffs, NJ: Merrill an imprint of Prentice Hall.

Nolen, S. B., & Nicholls, J. G. (1994). A place to begin again in research on student motivation: Teachers' beliefs. Teaching and teacher education, 10, 57-69.

Oosterhof, A. (1996). Developing and Using Classroom Assessments. Englewood Cliffs, NJ: Merrill, an imprint of Prentice Hall.

Rekase, M. (1995, December). Using portfolios for high stakes assessments. Presented at the Washington State Assessment Conference, Seattle, WA.

Resnick, L. B. & Resnick, D. P. (1991). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement, and instruction. Boston: Kluwer.

Salvia, J., & Ysseldyke, J. E. (1995). Assessment, Sixth Edition. Boston: Houghton Mifflin Company.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. Educational Measurement: Issues and Practice, 10 (1), 3-6, 12.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38 (3), 57-63.

Schon, D. A. (1987). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. SF: The Jossey-Bass Higher Education Series.

Schwab, J. J. (1978). Science, curriculum, and liberal education. Chicago: University of Chicago Press.

Shavelson, R. J. & Baxter, G. P. (1992, May). What we've learned about assessing hands-on science. Educational Leadership, 20-25.

Smith, M. L. (1991). Meanings of test preparation. American Educational Research Journal, 28 (3), 521-542.

Sommers, N (1982). Responding to student writing. College Composition and Communication, 33 (2), 148-156.

Stiggins, R. J. (1991). Relevant training for teachers in classroom assessment. Educational Measurement: Issues and Practice, 10 (1), 7-12.

Stiggins, R. J. (1994). Student centered classroom assessment.. New York: Merrill, an imprint of Macmillan College Publishing Company.

Stiggins, R. J., & Bridgeford, N. J. (1988). The ecology of classroom assessment. Journal of Educational Measurement, 22 (4), 271-286.

Stiggins, R. J., & Faires-Conklin, N. (1988). Teacher training in assessment. Portland, OR: Northwest Regional Educational Laboratory.

Stiggins, R. J., & Faires-Conklin, N. (1992). In teachers' hands: Investigating the practices of classroom assessment.. Albany, NY: SUNY Press.

Stiggins, R. J., Faires-Conklin, N., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice, 5 (2), 5-17.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. Journal of Educational Measurement, 26 (3), 233-246.

Stuck, I. (1995, April). Heresies of the new unified notion of test validity. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Taylor, C. S. (in press). Using portfolios to teach teachers about assessment: How to survive. Educational Assessment.

Taylor, C. S. & Nolen, S. B. (1996). A contextualized approach to teaching teachers about classroom-based assessment. Educational Psychologist, 31 (1), 77-88.

Toom, A. (1993). A Russian teacher in America. Journal of Mathematical Behavior, 12, 117-139.

Valencia, S. (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. Reading Teacher, 43 (4), 338-340.

Vidich, A. J., & Lyman, S. M. (1994). Qualitative methods: Their history in sociology and anthropology. In Handbook of Qualitative Research, Denzin, N. K. & Lincoln, Y. S. (Eds.). Thousand Oaks, CA: SAGE Publications, Inc., 23- 59.

Whyte, W. F. (1943). Street corner society: The social structure of an Italian slum. Chicago: University of Chicago Press.

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. Journal of Teacher Education, 42 (1), 37-42.

Wolf, D. P. (1991). Assessment as an episode of learning. In R. Bennett and W. Ward (Eds.), Construction versus choice in cognitive measurement. Hillsdale, NJ: Lawrence Erlbaum Associates.

Worthen, B. R., Borg, W. R., & White, K. R. (1993). Measurement and Evaluation in the Schools. New York: Longman.

## Appendix A
## Course evaluation items common across all evaluation forms

| Section | Item | Stem |
|---|---|---|
| 1: General Evaluation | 1 | Course as a whole |
| | 2 | Course content |
| | 3 | Instructor's contribution to the course |
| | 4 | Instructor's effectiveness in teaching the subject matter |
| 2: Feedback to Instructor | 1 | Course organization |
| | 3 | Explanations by instructor |
| | 4 | Instructor's ability to present alternative explanations |
| | 5 | Instructor's use of examples and illustrations |
| | 7 | Student confidence in instructor's knowledge |
| | 8 | Instructor's enthusiasm |
| | 11 | Availability of extra help when needed |
| 3: Information to Other Students | 1 | Use of class time |
| | 2 | Instructor's interest in whether students learned |
| | 3 | Amount you learned in the course |
| | 4 | Relevance and usefulness of course content |
| | 5 | Evaluative and grading techniques (tests, papers, projects) |
| | 6 | Reasonableness of assigned work |
| | 7 | Clarity of student responsibilities |

## Appendix B

Post-course survey items:

1. Please check the methods of assessment you are using in your field placement (list of 12 types of assessment, including worksheets, lab write-ups, observational records, paper-pencil tests, written reports, portfolios, peer evaluations)
2. Use the pie chart below to estimate the portion of your planning time you use each week to do the following activities (various planning activities, including planning lessons, assessments, units, writing objectives, etc.)
3. For each of the following situations, how often do you think about assessment issues? (3-point scale: frequently, sometimes, rarely); list of ten situations, including teaching, grading, planning instruction, observing other teachers, riding to and from work.
4. Thinking back on (the course) have any ideas or other aspects of the course influenced your teaching? If so, what part of (the course) has influenced your teaching the most? How has this influenced your teaching?
5. Have you had any new thoughts, questions, or understandings about assessment this quarter? If so, what are they?
6. Have you wrestled with any validity issues in your field placement this quarter? If so please describe one such issue.
7. Have you wrestled with any reliability issues in your field placement this quarter? If so please describe one such issue.
8. Have you taught all or part of the unit you designed for EDPSY 308? (For traditional course students: Have you used any of the materials or assessments you developed?)
9. If so, how helpful was the original plan or planning process? (5-point scale)

## About the Authors

### Catherine S. Taylor

Assistant Professor of Educational Psychology
312 Miller Hall, Box 353600
University of Washington
Seattle, WA 98195-3600

Voice phone: 206-543-1139
FAX: 206-543-8439

E-mail: ctaylor@u.washington.edu

EDUCATION

Ph.D. University of Kansas, 1986: Educational Psychology and Research
M.S.E. University of Kansas, 1978: Counseling Psychology
B.S.E. University of Kansas, 1974: Language Arts Education

EMPLOYMENT

1991- Assistant Professor, University of Washington, Educational Psychology-Research and Measurement
1986-1991 Senior Editor/Senior Project Manager, CTB/McGraw-Hill
1984-1986 Psychometrician, Psychological Corporation

RESEARCH INTERESTS

My main research focuses are large scale assessment development issues, validity theory, test theory, and research in the preparation of teachers. Current projects include studies of different scoring methods for performance-based tests in mathematics, reading, and writing, and a study of the philosophical foundations for and the social consequences of tests and testing practices.

**Susan Bobbitt Nolen**

Associate Professor of Educational Psychology
University of Washington
322 Miller Hall, Box 353600
University of Washington
Seattle, WA 98195-3600

Voice phone: 206-543-4011 ('96-'97 only)
206-543-1846
Fax: 206-543-8480 ('96-'97 only)
206-543-8439
sunolen@u.washington.edu

EDUCATION
Ph.D. Purdue University, 1986: Educational Psychology
M.Ed. Lewis & Clark College, 1976: Education of the Hearing-Impaired
B.A. Portland State University, 1975: Speech Pathology & Audiology

EMPLOYMENT

- 1990- Associate Professor, University of Washington Educational Psychology-Human Development & Cognition
- 1986-90 Assistant Professor, Arizona State University West Educational Psychology
- 1978-80 Teacher, Oregon School for the Deaf, Salem, OR High School English and Reading
- 1976-77 Teacher, Lacey Elementary School, Lacey, WA North Thurston Regional Program for the Hearing-Impaired

RESEARCH INTERESTS
My main research focus is the relationship between motivation and learning, and how this relationship develops over time. Current projects include investigations of how motivation develops differently depending on the learner's interpretation of

their social context for learning. A second interest is in assessment in schools, and the effects of various policies and practices on teacher and student motivation.

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://seamonkey.ed.asu.edu/

*Education Policy Analysis Archives* are "gophered" in the directory Campus-Wide Information at the gopher server INFO.ASU.EDU.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@sage.cc.purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*

Education Policy Analysis Archives

Volume 4 Number 19 December 24, 1996 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass Glass@ASU.EDU.

College of Education

Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

---

# Inclusive Education in the United States: Beliefs and Practices Among Middle School Principals and Teachers

**C. Kenneth Tanner**
**The University of Georgia**

**Deborah Jan Vaughn Linscott**
**Fulton County (GA) Schools**

**Susan Allan Galis**
**Commerce City (GA) Schools**

## Abstract

School reform issues addressing inclusive education were investigated in this nationwide (United States) study. A total of 714 randomly selected middle school principals and teachers responded to concerns about inclusion, "degree of change needed in" and "importance of" collaborative strategies of teaching, perceived barriers to inclusion, and supportive activities and concepts for inclusive education. There was disagreement among teachers and principals regarding some aspects of inclusive education and collaborative strategies. For example, principals and special education teachers were more positive about inclusive education than regular education teachers. Collaboration as an instructional strategy for "included" students was viewed as a high priority item. Responders who had taken two or more courses in school law rated the identified barriers to inclusive education higher than those with less formal training in the subject.

## Introduction to the Problem

The problem we addressed in this work was defined as a perceived lack of information about the issues surrounding inclusion (inclusive education) among middle school principals and teachers. We wanted to know the answer to the following question: What are the perceptions of front-line middle school educators regarding inclusion as a viable educational delivery system for students with disabilities? Background

The presentation of the April, 1983 report by the National Commission on Excellence in Education, A Nation at Risk, and other similar reports awakened Americans. These reports inaugurated the current waves of educational reform in the United States. Shapiro et al. (1993)

delivered a comparable wake-up call to the field of special education with their treatise "Separate and Unequal: How special education programs are cheating our children and costing taxpayers billions each year." Several issues were emphasized. For example, labels and categorizations varied from state to state.

Schiller, Countinho, and Kaufman (1993) insisted that educational reform and restructuring initiatives require special education to be united with regular education. A few of the demands placed on general education were to provide inclusion for students with disabilities through the Regular Education Initiative (REI) and to provide a sophisticated work force for the 21st Century. Repositioning of special education includes policies for the integration of students with disabilities (Wade and Moore, 1992). In contrast, segregated programming emphasizes differences while promoting dependence and decreasing self-sufficiency (Byrnes, 1990).

Poignant debate has materialized over the re authorization of The Individuals with Disabilities Education Act (IDEA), the 1990 re authorization of the original P.L. 94-142. The current re authorization for IDEA has experienced delays, extensions, and debate in and out of the field of special education. One area of impassioned or "thorny" discussion has been the requirements for a free appropriate public education (FAPE) in IDEA and the preference for mainstreaming "embodied in federal special education law" (Huefner, 1994, p. 27).

If the law has been massively successful in assigning responsibility for students, it has been less successful in removing barriers between general and special education. It did not anticipate that the artifice of delivery systems in schools might drive the maintenance of separate services and keep students from that mainstream, or that the resources to fund these services would be constrained by economic forces (Walker, 1987).

The National Council on Disability (1995) reported to the United States President on the re authorization of IDEA. The issue of least restrictive environment (LRE) was one of the ten basic themes addressed both historically and as a current theme in the re authorization of IDEA. The Council concluded that the re authorization must be pursued and that it should address the improved implementation of IDEA. "The Court has made it clear that IDEA is not one of the so-called "unfunded Federal mandates," but is a Federal grant program that is entirely justified under Congress' power . . . More than that, the Court has acknowledged in the most unequivocal terms that IDEA provides Federal aide to the States to help them carry out their own legal obligations to educate all children, including those with disabilities." (p. 4)

The decision in Smith v. Robinson (1984) underscored this: "Congress made clear that the [IDEA] is not simply a funding statute. The responsibility for providing the required education remains on the States. . . . And the Act established an enforceable substantive right to a free appropriate public education" (p. 1009-1010).

While "inclusion" is not a term used in the law and regulations, it is currently the often used terminology to indicate consideration of the least restrictive environment for students with disabilities. The statute defined the consideration of least restrictive environment as:

> . . . procedures to assure that, to the maximum extent appropriate, children with disabilities, including children in public or private institutions or other care facilities, are educated with children who are not disabled, and that special classes, separate schooling, or other removal of children with disabilities from the regular educational environment occurs only when the nature or severity of the supplementary aids and services cannot be achieved satisfactorily. ([IDEA] $1412 [5][B][1990])

Opponents of inclusion have emphasized the need to maintain a full continuum of services and argued that those expounding "full" inclusion had overlooked this provision of the IDEA. Vergason and Anderegg (1992, 1993) argued that an inclusive classroom was not in the "least restrictive environment" interests of most students with disabilities. Fuchs and Fuchs (1994) identified The Association for Persons with Severe Handicaps (TASH) as the leader in the reform movement for inclusion, and warned that TASH did not speak for all groups in their desire for full inclusion, but that ". . . their continued provocative rhetoric will polarize a field already agitated." (p. 305)

## Conceptual Basis for the Study

Baker, Wang, and Walberg (1995) traced the beginnings of inclusion to a report by Heller, Holtzman, and Messick through the National Academy of Sciences in 1982. The panel of Heller et al. found the classification and placement of children in special education ineffective and discriminatory. A comparison of the effects of inclusive versus non inclusive educational practices for special education students has been made by Baker (1994), Carlberg and Kavale (1980), and Wang and Baker (1985). A meta-analysis demonstrated a "small-to-moderate beneficial effect of inclusive education on the academic and social outcomes of special needs children" (Baker et al., 1995, p. 34). Baker et al. asserted that the "concern is not whether to provide inclusive education, but how to implement inclusive education in ways that are both feasible and effective in ensuring school success for all children, especially those with special needs." (p. 34)

According to Yatvin (1995), side effects of the resource pull-out program have enhanced the idea of inclusion. Many drawbacks of the resource pull-out program model have been underscored: special education resource rooms often served 12 to 15 diverse students, students brought a variety of needs from several different grade levels, the special education teacher gave very little active instruction, and instruction occurring was skill related and not tied to classroom themes.

The outcomes for non disabled students in classes with included disabled peers had been identified as a barrier to inclusion. Available research revealed no statistically significant effects on the academic outcomes of the non disabled peers (Staub & Peck, 1995). Instructional time was not lost by non disabled students when disabled students were included in their classrooms. Additionally, non disabled peers did not pick up undesirable behaviors from their disabled peers.

Parents and teachers of non disabled peers in an inclusive setting reported no developmental harm to the children (Bailey & Winston, 1989; Giangreco et al.,1993; Green & Stoneman, 1989; and Peck et al., 1992). Helmstetter, Peck and Giangreco (1993) surveyed non disabled students who were in inclusive high school settings. The non disabled peers reported that they had not missed out on any valuable experiences because of their inclusive experience.

Five positive outcomes for non disabled peers were identified by Staub and Peck (1995): reduced fear of human differences accompanied by increased comfort and awareness, growth in social cognition, improvements in self-concept, development of personal principles, and warm and caring friendships (p. 37-38). The literature from the review of research on non-disabled peers pointed to inclusion as a positive experience for both non disabled and disabled students, helping to build a basis for community and friendships.

Yatvin identified a major factor that led to the philosophy of inclusion: "All children learn best in regular classrooms when there are flexible organizational and instructional patterns in place and human and material supports for those with special needs." (p. 484) Sapon-Shevin (O'Neil, 1995) used the current "politically correct" rhetoric in explaining the basis of a philosophy for inclusion: "As far as a rationale, we should not have to defend inclusion -- we should make others defend exclusion. There's very little evidence that some children need segregated settings in which to be educated. At another level, we know that the world is an inclusive community. . . . So we should begin with the assumptions that all children are included and that we must meet their needs within an inclusive setting." (p. 7)

Van Dyke, Stallings and Colley (1995) identified fundamental arguments to support the philosophy of inclusion. One major argument was that segregating the students classified them, created bias, and made them different. They were set apart from the classroom community.

Stainback and Stainback (1984) proposed a merger of regular and special education into one unified system. This assertion was based on two premises: the instructional needs of students did not warrant a dual system, and the operation of a dual system was viewed as inefficient. Others in the field of special education (Hobbs, 1980; Meyen, 1978; Reynolds & Birch, 1982; Ysseldyke & Algozine, 1982) had set the stage for Stainback and Stainback to assert the merger of special and general education as the next natural step in the evolution of education for students with disabilities. Sapon-Shevin (1990), suggests that academic and functional skills can be met in the regular classroom setting. Reynolds and Birch (1982) stated that "the whole history of education for exceptional students can be told in terms of one steady trend that can be described as progressive inclusion" (p. 27).

Fuchs and Fuchs (1995) compiled information from four major efficacy studies and found that "for certain students, special education programs appear to promote greater academic achievement than do regular classrooms" (p. 526). Research concerning the beliefs and practices of middle school personnel regarding inclusion was scarce (Farley, 1991; Rath, 1989; White, 1993). The available research was regional in nature, confined to a single state or a single school district.

## Context of the Study

This steady trend toward inclusion invited investigation of middle school educators. The front-line educators were studied concerning their agreement with the assertions that students with disabilities could benefit from instruction in the regular education classroom. The current climate underscored the need for answers to questions about inclusion from the professionals who were the providers of service. Their (key players) viewpoints needed to be identified and documented.

We made the assumption that it is important to gather information from people who have the responsibility to implement inclusion. We contend that their experience and insight is vital in shaping future educational trends for all students.

Many advocates of school reform assumed that support existed for inclusion among those educators who would be the primary change agents -- the principals, general education teachers, and special education teachers. Little data existed to support this, and the number of critics matched supporters in the literature. Teacher unions and many general education professional organizations voiced opposition to inclusion. Consequently, we viewed this study as a robust procedure to generate information about the beliefs and practices of middle school personnel representing various schools and groups across America.

McDonnell and Hardman (1989) examined the role of all school personnel in the desegregation of students with disabilities. They designated regular education principals as key players in the quality of special education services and the degree of successful integration efforts and concluded that the attitudes of the principals appear to be even more important than their actions.

The literature on the role of the principal in effecting needed modifications to accommodate inclusion offered some insights into the process of change. Riley (1993) underscored the role of the building level principal and teachers in any change process and the need for input from them into proposed changes: "I've learned . . . that the bottom-up approach works when you involve the nuts-and-bolts people. Who knows better than site school administrators and teachers the kind of changes that have the best chance of improving education?" (p. 5) Burrello (1991) stated that effective principals make no distinction between the expectations set for special and general education students, staff, and programs.

Middle schools have traditionally been organized differently than elementary schools with the delivery of services centered around team approaches. The impact of inclusion on these structures might be expected to produce a new and different set of challenges than those presented in the elementary schools. Given these circumstances, we concluded that investigations of middle school personnel and the resulting beliefs and practices in relation to inclusionary practices would be an addition to this sensitive body of knowledge.

## Purpose of the Study

The purpose of this study was to investigate the beliefs and practices of a national sample of middle school personnel (principals, general education teachers, and special education teachers). We designed a survey that provided an avenue to question those who directly implement policies and procedures of school reform issues influencing the delivery of services to students with disabilities. Demographic and career information were contrasted with responses to ascertain if significant differences among the variables existed.

This inquiry paralleled the work of Galis and Tanner who investigated elementary school principals, special education administrators, and teachers in the schools of the state of Georgia. It was undertaken to broaden the application of Galis' survey instrument by studying a special database (Galis, 1994; Galis & Tanner, 1995). MacKinnon and Brown (1994) reported that secondary schools "in part because of the historical-structural characteristics of these organizations, embody different and perhaps more complex problems [than elementary schools] in meeting the demands of inclusive educational practices" (p. 126). Anderman and Maehr (1994) argued that student motivation differed in middle school from elementary school settings. Students generally receive instruction through a team delivery system at the middle school level while elementary schools traditionally deliver services through self-contained classrooms.

Given the arguments found in the literature and research, we defined the dependent variables as inclusive education, collaborative strategies, perceived barriers to inclusion, and supportive activities and concepts for inclusive education. Independent variables were the current role of the respondent, number of years in current position, number of years as a school administrator, number of years in education, and the number of courses taken in school law.

### Variables
### Inclusive Education
### Instructional Strategies

Several studies (Madden, Slavin, Karweit, Dolan, & Wasik, 1993; Slavin et al.,1991; Slavin, Madden, Karweit, Livermon, & Dolan, 1990) have pointed to individualized instruction, cooperative and peer mediated instruction, and teacher consultation models as programs that would support teachers in their attempts to fully integrate academically students with disabilities.

Jones and Carlier (1995) reported that middle school students with multiple disabilities were successfully included in a collaborative setting using cooperative learning activities. Original goals for the students with disabilities were to increase the time spent in the general education classroom and to improve the quality of functional instruction given while in the general education classroom. Peer and teacher interactions increased for learners with disabilities. Special education teachers reported having a better perception of appropriate grade-level behavioral and academic expectations. Non disabled students shared their observations of the likenesses between themselves and the students with disabilities. The non disabled students were sharing tasks and adapting jobs so the students with disabilities were participants rather than just observers.

Jenkins et al. (1994) studied an approach combining Cooperative Integrated Reading and Composition (CIRC), cross-age tutoring, supplementary instruction in synthetic phonics, and in-class instructional support from specialists. Regular, special education, and Chapter I students showed significantly improved scores in the experimental group, as measured by the Metropolitan Achievement Test, in reading vocabulary, total reading, and language, with marginally significant gains in reading comprehension.

In another study, students with learning disabilities served through resource programs one period daily were compared to those served through consultative services combined with in-class instruction and consultative services to the teachers. Analysis of student achievement scores showed that students receiving a combination of consultative and direct services exhibited small, but significantly greater overall gains in achievement than did students receiving resource intervention one period daily (Schulte, Osborne, & McKinney, 1990).

### Principals, Regular Educators, and Special Educators

A National Association of Elementary School Principal's poll (Principals favor reconsideration, 1995) indicated that responding principals were not in support of "full inclusion." Twenty-seven percent agreed with the premise that all children should be assigned to regular classes despite disability, 72% disagreed and 1% had no opinion. The executive director of the association summarized: "Children learn an enormous amount from each other that they can't learn from teachers or parents and the great majority of disabled youngsters benefit socially, psychologically and academically from joining their peers in regular classrooms. . . . But the concept of inclusion has been pushed to such extremes that it's robbing non-handicapped children of their right to learn, while depriving handicapped children of the specialized teaching they need." (p. 2)

Burrello and Wright (1992) identified effective practices of principals who had participated in programming for the inclusion of students with disabilities. Two important practices noted were to provide opportunities for the faculty and staff to discuss integration in light of consensus values and belief statements; and create a special support group of faculty and staff for the purpose of brainstorming and facilitating integration, mainstreaming, and inclusion efforts.

Farley (1991) studied middle school personnel in Virginia and found attitudes toward the integration of students with disabilities similar to attitudes of personnel in other grade levels. Principals had more favorable attitudes than teachers toward the integration of students with disabilities. Factors found significant concerning the attitudes of personnel were prior experience working with persons with disabilities, educational background, and course work in special education.

Baines, Baines, and Masterson (1994) documented the frustration of teachers in a middle school who were meeting the needs of students with disabilities in the regular education classrooms without the support needed for the student, the teacher, and the other classmates. All teachers except the physical education teacher reported heightened stress due to mainstreaming and 20% of the respondents on a school-wide survey reported that they were reconsidering teaching as a career.

Raison, Hanson, Hall, and Reynolds (1995) indicated that the problems that Baines et al. (1994) had encountered were not due to mainstreaming, but to "inadequate communication, misgovernance and poor allocation of resources." (p. 481)

Schumm and Vaughn (1992) studied 775 teachers representing 39 schools in a metropolitan school district in the Southeast. Elementary teachers were more likely to make adaptations in preplanning, interactive planning, and post planning. Planning for mainstreamed students was frequently inhibited by class size, lack of teacher preparation, problems with emotionally handicapped students, and limited instructional time.

### Collaborative Strategies

The collaborative team approach has emerged as a model of addressing the curricular needs of all children, both disabled and non- disabled in the same classroom (Nevin, Thousand, Paolucci-Whitcomb & Villa, 1990; Villa & Thousand, 1992). In the Supportive Teaching Model (Bauwens, Hourcade, & Friend, 1989), general education teachers are responsible for the content of the material, while the special educator accepts responsibility for the adaptations. Material presentation, follow- up, lecture and other methods are cooperatively planned and presented. The Co-teaching or Team-teaching Model incorporates shared planning, instruction, and monitoring of performance and evaluations. Regular and special education teachers are equals in the classroom. The Complementary Model uses the special educator to weave techniques and strategies into the general education curriculum.

Lipsky (1994) reported that a survey by The National Center on Educational Restructuring and Inclusion (NCERI) indicated there were several models of inclusive education based on differing teacher roles: Co-Teaching Model, Parallel Teaching (the special education teacher works with a small group of special education students in an area of the general education classroom), Co-Teaching Consultant Model (the special education teacher operates both a pull-out and a co-teaching arrangement), Team Model (the teaming of special and general education teachers who accept the responsibility for all students, including those with disabilities), and Methods and Resources Teacher Model (the special education teacher works with the general education teachers as a resource person).

The literature is rich with works on collaborative teaching. For example, Thousand and Villa (1992) reviewed needed aspects of collaborative teams and the dynamics they add to restructuring; West and Cannon (1988) examined competencies needed for effective collaborative strategies for special and regular educators; Maroldo (1994) found that special and general education teachers needed to learn a common language, due to the isolation they have experienced; and Detmer, Thurston, and Dyke (1993) authored a manual for collaboration in schools serving students with disabilities through collaborative teaching.

**Perceived Barriers to Inclusive Schooling**

The National Council on Disability (1995) explored barriers which could impede the implementation of identified promising practices in special education. One major barrier to the practice of inclusion is the reactive instead of proactive response of schools to students' special needs. Too often students are simply excluded, instead of school personnel working to overcome challenging behaviors. Another barrier hinges on the fact that some schools still do not make the environmental modifications that would increase access. A third and attitudinal barrier concerns general educators' lack of feeling responsible for educating students with disabilities.

Hasazi, Johnston, Liggett, and Schattman (1994) conducted a multistate, qualitative study of the LRE provision of the IDEA, 1989 to 1992. Six facets seemed to influence the implementation of LRE: finance, organization, advocacy, implementors, knowledge and values, and state/local context. Possible barriers to inclusion were student outcomes, policy and bureaucracy concerns, staff development and training, funding issues, and the stand of some professional organizations. Supportive activities and concepts for inclusive education

Many practices reported as helpful or supportive to inclusionary factors were the inverse of the factors reported in the prior section addressing barriers. The National Council on Disability list of barriers (1995) could be stated in positive terms as supports to inclusion.

The National Center on Educational Restructuring and Inclusion (NCERI, 1994) at City University of New York reported six classroom practices which had allowed inclusion to succeed: multi-level instruction, cooperative learning, activity-based learning, mastery learning, technology, and peer support and tutoring programs (Lipsky, 1994, p. 5). Other factors determined to be "necessary for inclusion to succeed" were: visionary leadership, collaboration, refocused use of assessment, supports for staff and students, funding, and effective parental involvement (p. 5-7).

Schools in Newark, Delaware were reported to have included children in regular education classrooms for the past twenty years (Johnston, Proctor, & Corey, 1995). Their Team Approach to Mastery (TAM) project resulted in a school district of 20,000 students functioning without any resource classrooms. One hundred TAM classrooms serve special education students in a general education environment. TAM's successes were attributed to seven factors: team teaching, learning centers, ego groups, direct instruction, positive approach, point cards, and teacher cadres. TAM's approach offers children "not a way out of general education, but a way in." (p. 47)

General and special education elementary teachers (N=158) who had been involved in inclusive education were surveyed concerning their perceptions of supportive practices for inclusion (Wolery, Werts, Caldwell, Snyder, & Lisowski, 1995). One major finding was that special and general educators reported similar levels of need for resources, but special educators reported greater availability of resources than general educators. A high percentage of respondents reported a need for training and a low percentage reported having training. Research Questions

The research questions were based on the gaps in the research and literature and our interests that were sparked by experience. Based on the assumption of "lack of information regarding inclusive education in middle schools", the context of the variables, and the conceptual background, four research questions were formed: Is there a statistically significant difference among the independent variables regarding the beliefs and practices of middle school personnel when considering

1. inclusive education,
2. collaborative strategies,
3. factors perceived as barriers to inclusive education, and
4. supportive activities and concepts for inclusive education?

**Method**

**Research Design**

Schools were selected randomly. A sample was drawn from all middle grade schools in the United States. The list of schools was purchased from the National Association of Secondary School Principals (NASSP) and only public school personnel were surveyed. The sample was selected from the population of 12,941 public middle and junior high schools. The error range for the sample was 4% ($d < .04$). Based on the observations of Gallup (1976, p. 69), a "confidence level of 95% and an error range of four percentage points are used by most survey agencies including the Gallup Poll." The sample size was calculated by using Nunnery's and Kimbrough's (1971) method of sampling. A sample of 574 schools was drawn from the population.

### Instrumentation

With the written permission of Galis (1994), selected questions from her questionnaire along with questionnaire items generated according to the conditions presented above were used to collect data for the study. The instrument focused on the beliefs and practices of middle school personnel (See Table 3 for questionnaire items).

Validity. The questionnaire was reviewed by a panel of experts including selected special education administrators to establish face and content validity. Suggestions for improvement were then incorporated. The wording was changed on some items as a result of the review. A pilot was completed and two items were challenged by the panel. These questions were deleted.

Reliability (Phase I). The reliability of the instrument was determined in two phases. Prior to dissemination, twenty (20) educators similar to the sample group were asked to volunteer to respond to the instrument. Two weeks later they responded to the same instrument again.

The items were then examined by using the repeated measure design. The t test for correlated sample means was used to test the null hypothesis of no significant difference between the two response probes for each question. The test-retest analysis had the decision criteria that Items exceeding the critical t value of 2.093 were to be removed from the instrument (Alpha = .05, df = 19). No items exceeded this value, so none were deleted from the instrument on that basis.

Reliability (Phase II). Data from the larger sample were analyzed according to Cronbach's alpha coefficient test to determine the reliability of the subsets. This test determined the correlation coefficient between the response to a single item and the response to other items in the subset. De Vaus (1986) designated an alpha coefficient of .70 as desirable. Items were removed if the omission of that item improved the subset alpha to .70 or higher. Consequently, item number 42 (variable 57) was removed. Coefficients for the five categories of dependent variables were: Inclusive education (.78), degree of change needed to include collaborative strategies (.82), importance of factors supporting integration of students (.71), factors perceived as barriers to an inclusive environment (.77), and factors perceived as supportive of an inclusive environment (.72).

### Constraints of the Study

This study addressed personnel at the middle school level. Results may not necessarily represent the beliefs and practices of personnel at the elementary and high school level.

This instrument was sent by U.S. Mail and some recipients may not have felt compelled to respond. Non responses may imply certain important issues that are not included in the study. Opinions may be used to infer or estimate the attitude of the respondent. Overt actions may be unrelated to the actual attitude of the individual (Best,1970).

### Data Collection

A packet of three sets of surveys was mailed to the principal of each school. The principal was requested to fill out one questionnaire and distribute the other questionnaires to the first general education teacher on the school roster and the first special education teacher on the school roster. A cover letter explained the purpose of the study and gave instructions for distribution. Each questionnaire was in a booklet form such that the respondents could staple it closed for mailing. Questionnaires were pre-stamped and addressed. Respondents were offered a copy of the summary of the results of the study. A stamped postal card addressed to the investigator was enclosed for each of the participants to mail separately. This separate medium helped to preserve the anonymity of the respondents and possibly serve as an incentive to respond to the survey instrument. A statement to be checked on the postcard stated: "Yes, I have completed and mailed the questionnaire and would like to receive a summary of the results of this study." The respondents then printed their name with a preferred mailing address to receive a summary of the study results. The data collection began in November, 1994, and concluded in February, 1995.

**Data Presentation and Analysis**

Each variable was analyzed by frequency of response and comparisons were also made among the variables. Both one-way and two- way analysis of variances (ANOVA) were generated (Alpha = .05 ). Descriptive Data

Mailings to 574 schools included 1722 questionnaires. The response rate was 41.5% and consisted of 714 returns. Table 1 indicates the results of the responses to the independent variables. Thirty-six and seven-tenths percent of the responders was in the principalship role (n = 262), 31.6 percent reported that they were regular education teachers (n =228), and 31.4 percent of the responders taught special education. The variable for years in current position was divided into 1-2 years, 3-5 years, 6-10 years, and 11-37 years groupings to approximate 25% in each category. One hundred seven respondents reported they had taken more than two courses in school law. Table 2 presents general demographic information.

### Table 1
### Descriptors for the Six Independent Variables

| Independent Variable | Descriptors | Percentage* | N |
|---|---|---|---|
| Current position | Principal | 36.7% | 262 |
| | General Ed Teacher | 31.9% | 228 |
| | Special Ed teacher | 31.4% | 224 |
| Number of years in current position | 1-2 years | 24.8% | 173 |
| | 3-5 years | 25.8% | 180 |
| | 6-10 years | 26,2% | 182 |
| | 11-37 years | 23.1% | 161 |
| Number of years in education profession | 1-12 years | 25.8% | 180 |
| | 13-19 years | 24.0% | 167 |
| | 20-24 years | 24.8% | 173 |
| | 25-42 years | 25.4% | 177 |
| Courses in school law | 1 course | 46.8% | 225 |
| | 2 courses | 31.0% | 149 |
| | More than 2 courses | 22.2% | 107 |
| Years as a school administrator | 1-6 years | 24.3% | 63 |
| | 7-10 years | 26.7% | 69 |
| | 11-16 years | 23.5% | 61 |
| | 17+ years | 25.5% | 66 |

*Missing cases were excluded.

### Table 2
### Demographic Data for Respondents

| | Mean | N | St. Dev. | Range |
|---|---|---|---|---|
| Years in Current Position | 7.3 | 696 | 6.30 | 1-37 |
| Total Years in Education | 18.39 | 697 | 8.56 | 1-42 |
| Courses in School Law | 2.06 | 481 | 1.90 | 1-20 |

**Items and Subsets**

The individual item means and standard deviations for all respondents by cluster of questions per dependent variable are shown in Table 3 in the Appendix. Item to variable position is indicated. The first question in Section II was variable 16, since the first 15 variables were demographic. Item means ranged from 5.515 (highest) to 1.999 (lowest). Item 39 (importance of collaboration) had the highest mean for all items.

**Findings**

Both one-way and two-way ANOVAs were used to study the mean differences among the groups. The Scheffe' test was applied to determine where statistically significant differences existed among the subgroups (Alpha = .05).

**Research Question One**

Is there a statistically significant difference among the independent variables regarding inclusive education? Items 1 through 12 in Table 3 deal with the subset on inclusive education.

There was a significant difference regarding inclusive education by position (F = 19.63, p = .001). The Scheffe' analysis revealed that principals (mean = 4.54) and special education teachers (mean = 4.59) more strongly agreed with the statements about inclusive education than did regular education teachers (mean = 4.16). Principals' and special education teachers' mean responses were significantly higher than those of regular education teachers (Table 4). Special education teachers' mean responses were significantly different from regular education teachers. No other significant differences were found among the variables when compared to the "inclusive education category."

### Table 4
### Inclusive Education by Position

| Source | df | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 24.10 | 12.05 | 19.63 | .001 |
| Within Groups | 694 | 426.10 | .61 | | |
| Total | 696 | 450.19 | | | |

| Group | Count | Mean | Standard Deviation | Standard Error |
|---|---|---|---|---|
| Principals | 258 | 4.54 | .712 | .043 |
| Reg Ed Tchers | 221 | 4.16 | .921 | .062 |
| Spec Ed Tchers | 218 | 4.59 | .710 | .048 |
| Total | 697 | 4.44 | .804 | .031 |

---

**Research Question Two**

Is there a statistically significant difference among the independent variables regarding collaborative strategies? Questionnaire items 13-15 addressed the degree of change needed regarding collaborative strategies; and items 39, 40, 41, and 43 measured the perceived importance of integrating students with disabilities into general education settings (See Table 3).

A statistically significant relationship existed among collaborative strategies by position for both components. For example, Table 5 shows that a statistically significant difference existed between regular education teachers and special education teachers on "the need for change" (F = 4.11, p = .017). According to the post hoc test, regular education teachers' mean response (4.79) were significantly lower than special education teachers' mean response (5.05). There was no statistically significant difference between principals (4.97) and teachers' perceptions.

Table 6 displays a statistically significant difference in the perceived importance of collaborative strategies when compared by position (F = 4.67, p = .010). Both principals and special education teachers had significantly different perceptions than regular education teachers as determined by the Scheffe' test. Regular education teachers perceived integration of students to be less important than the other two groups.

## Table 5
## Degree of Change Needed in Education (Collaboration)

| Source | df | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 7.53 | 3.76 | 4.11 | .017 |
| Within Groups | 700 | 641.89 | .92 | | |
| Total | 702 | 649.42 | | | |

| Group | Count | Mean | Standard Deviation | Standard Error |
|---|---|---|---|---|
| Principals | 261 | 4.97 | .887 | .055 |
| Reg Ed Tchers | 221 | 4.79 | 1.049 | .071 |
| Spec Ed Tchers | 221 | 5.05 | .943 | .063 |
| Total | 703 | 4.94 | .962 | .036 |

## Table 6
## Importance of Collaboration

| Source | df | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 5.23 | 2.62 | 4.67 | .010 |
| Within Groups | 697 | 390.26 | .56 | | |
| Total | 699 | 395.49 | | | |

| Group | Count | Mean | Standard Deviation | Standard Error |
|---|---|---|---|---|
| Principals | 260 | 5.30 | .669 | .042 |
| Reg Ed Tchers | 223 | 5.13 | .905 | .061 |
| Spec Ed Tchers | 217 | 5.33 | .655 | .045 |
| Total | 700 | 5.25 | .752 | .028 |

No significant differences were found when the number of years in the respondent's current role was compared to the items concerning collaborative strategies. A statistical significance (F = 3.74, p = .011) was found for items pertaining to perceived importance of collaborative strategies when compared to total years of educational experience. The post hoc analysis revealed that those persons in group two (13 through 19 years in education) scored significantly higher than respondents in group one (1 through 12 years). This parallels the Galis and Tanner (1995) findings that show younger teachers to be less open to new ideas. Results are presented in Table 7. Years in administrative positions for principals were analyzed and no significant results were identified. No significant relationship was identified when collaborative strategies were compared to the number of courses taken in school law.

## Table 7
## Importance of Collaboration

| Source | df | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 6.27 | 2.09 | 3.74 | .011 |
| Within Groups | 679 | 379.54 | .56 | | |
| Total | 682 | 385.81 | | | |

| Group | Count | Mean | Standard Deviation | Standard Error |
|---|---|---|---|---|
| Group 1 (1-12 yrs) | 178 | 5.15 | .776 | .058 |
| Group 2 (13-19 yrs | 163 | 5.38 | .767 | .060 |
| Group 3 (20-24 yrs) | 168 | 5.16 | .760 | .059 |
| Group 4 (25-42 yrs) | 174 | 5.30 | .686 | .052 |
| Total | 683 | 5.25 | .752 | .029 |

### Research Question Three

Is there a statistically significant difference among the independent variables regarding factors perceived as barriers to inclusive education? Items 16-28 pertained to barriers (See Table 3).

According to the analysis of variance test, the responder's position was not a statistically significant factor to be considered as barriers to inclusion. Years of experience in current position, total years in education, the number of years of administrative experience for principals, and total years of education experience for principals did not yield significant results regarding barriers.

Responses to the items about barriers and the number of courses taken in school law were analyzed and a statistically significant relationship was established (F = 3.45, p = .032). Data are presented in Table 8. The Scheffe' analysis revealed a significant difference between Group 2 (those who took 2 law courses) and the other two groups. Group 2 showed the strongest agreement with the statements about barriers.

A two-way ANOVA was completed for barriers by position by the number of school law courses taken. A statistically significance interaction (F = 2.629, p = .034) was identified (Table 9). There was a significant difference between the perceptions of principals and teachers. Principals reported lower mean responses to perceived barriers. Two or more school law courses appeared to explain the respondents' significant differences found regarding barriers in this two-way analysis. Figure 1 reveals the interaction between the number of school law courses and responder's position on perceived barriers.

### Table 8
### Barriers to Inclusive Education by Courses taken in School Law

| Source | df | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 3.97 | 1.99 | 3.45 | .032 |
| Within Groups | 438 | 251.74 | .57 | | |
| Total | 440 | 255.71 | | | |

| Group | Count | Mean | Standard Deviation | Standard Error |
|---|---|---|---|---|
| Group 1 (one course) | 209 | 3.13 | .725 | .050 |
| Group 2 | 131 | 3.35 | .793 | .069 |
| Group 3 (> 2 courses) | 101 | 3.23 | .780 | .078 |
| Total | 441 | 3.13 | .762 | .036 |

### Table 9
### Barriers by Position by School Law (2-Way)

| Source | df | Sum of Squares | Mean Squares | F | Prob. of F |
|---|---|---|---|---|---|
| Main Effects | 4 | 6.72 | 1.68 | 2.982 | .019 |
| Position | 2 | 2.74 | 1.37 | 2.437 | .089 |
| School law | 2 | 5.04 | 2.52 | 4.478 | .012 |
| 2-Way Interactions Position Schl Law | 4 | 5.20 | 1.48 | 2.629 | .034 |
| Explained | 8 | 12.63 | 1.58 | 2.806 | .005 |
| Residual | 432 | 243.08 | .56 | | |
| Total | 440 | 255.71 | .58 | | |

| | Cell Means / (n) | | |
|---|---|---|---|
| | Courses in School Law | | |
| Position | One | Two | Two or More |
| Principal | 3.18<br>(82) | 3.17<br>(76) | 3.14<br>(71) |
| Reg. Ed. Teacher | 3.04<br>(63) | 3.62<br>(28) | 3.46<br>(8) |
| Se. Ed. Teacher | 3.15<br>(64) | 3.58<br>(27) | 3.43<br>(22) |

N= 441; Mean =3.22

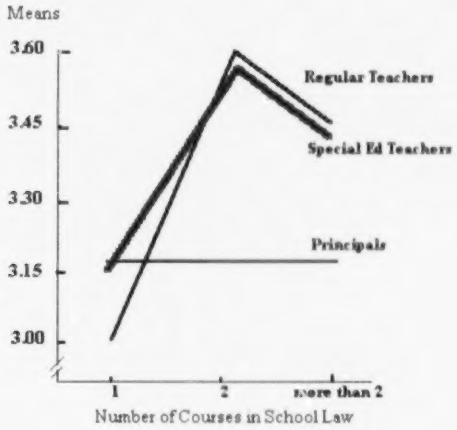

**Figure 1. Interaction between number of school law courses and position on perceived barriers.**

Table 10 presents the data about barriers ranked from the highest to lowest means. The top three perceived barriers were identified as lack of adequate staff size, lack of shared special/education planning time, and lack of amount of planning time allocated. School climate, negotiations with teachers organizations, and school board policy received the lowest rankings.

## Table 10
## Perceived Barriers to Inclusion

| Rank/ Item# | Variable | Mean | Standard Deviation | Descriptor |
|---|---|---|---|---|
| 1/19 | 34 | 4.503 | 1.512 | Lack of adequate size staff |
| 2/24 | 39 | 4.419 | 1.617 | Lack of shared planning |
| 3/23 | 38 | 4.291 | 1.568 | Not enough plan time |
| 4/17 | 32 | 3.794 | 1.432 | Confusion about roles |
| 5/21 | 36 | 3.605 | 1.546 | Lack of staff willingness |
| 6/18 | 33 | 3.420 | 1.647 | Federal rules/regulations |
| 7/16 | 31 | 3.280 | 1.464 | Concern: student outcomes |
| 8/28 | 43 | 2.986 | 1.640 | Weighted funding |
| 9/20 | 35 | 2.846 | 1.623 | Lack central office support |
| 10/27 | 42 | 2.761 | 1.607 | State rules and regs |
| 11/26 | 41 | 2.473 | 1.447 | School climate |
| 12/22 | 37 | 2.095 | 1.390 | Teacher unions |
| 13/25 | 40 | 1.999 | 1.325 | School board policy |

### Research Question Four

Is there a statistically significant difference among the independent variables regarding factors perceived as helpful or supportive of inclusive education? The items addressed in the questionnaire as possible supports to inclusion were 30-37 (See Table 3).

One statistically significant difference was found for this question. The data analysis for principals revealed a significance in the years of administrative experience related to perceived supports for inclusion ($F = 3.37$, $p = .019$). Group One (with one through six years of administrative experience) showed the strongest agreement with the perceived supports to inclusion. Group one (principals with at least six years experience) had a significantly higher mean than group four (17 -32 years). These data are presented in Table 11.

Table 12 presents variables perceived to be helpful and supportive of inclusion as ranked by the mean. The top three selections were clustered closely together: funds for staff training, funds and/or release time for staff collaborative planning, and a lead teacher trained in special education and instructional strategies. The choice with the lowest mean score was for an extra assistant principal who is a generalist.

## Table 11
## Perceived Supports to Inclusion for Principals by Years of Administrative Experience

| Source | df | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 7.36 | 2.45 | 3.37 | .019 |
| Within Groups | 246 | 179.08 | .73 | | |
| Total | 249 | 186.44 | | | |

| Group | Count | Mean | Standard Deviation | Standard Error |
|---|---|---|---|---|
| Group 1 (1-6 yrs) | 59 | 4.47 | .786 | .102 |
| Group 2 (7-10 yrs) | 67 | 4.16 | .801 | .098 |
| Group 3 (11-16 yrs) | 58 | 4.38 | .925 | .121 |
| Group 4 (17-32 yrs) | 66 | 4.03 | .897 | .110 |
| Total | 250 | 4.25 | .865 | .055 |

---

### Table 12
### Factors Perceived to be Helpful and Supportive of Inclusion

| Rank/Item | Variable | Mean | Standard Deviation | Descriptor |
|---|---|---|---|---|
| 1/35 | 50 | 5.280 | .976 | Funds for staff training |
| 2/34 | 49 | 5.250 | 1.070 | Funds/ release time for collaborative training |
| 3/36 | 51 | 4.994 | 1.174 | Lead teacher |
| 4/37 | 52 | 4.224 | 1.679 | School board support |
| 5/32 | 47 | 4.187 | 1.502 | De-emphasis test scores |
| 6/31 | 46 | 3.903 | 1.725 | Central office support |
| 7/33 | 48 | 3.621 | 1.792 | Flat funding formula |
| 8/30 | 45 | 3.154 | 1.802 | Extra assistant principal |

**Discussion of the Findings**

There was a significant difference found for current position of the respondents for the inclusive education and both collaborative strategies questions. A statistically significant difference was found for total years in education when compared to the importance of collaborative strategies variable. The number of school law courses was statistically significant for barriers to inclusion.

Arrington's study (1992) supported the current finding that years of educational experience were not significant in respondents' support for inclusive education. Principals and special education teachers were each significantly different from regular education teachers concerning their perceptions of inclusive education. Regular education teachers were significantly less supportive of inclusive education than the other two groups. Arrington (1993) and Farley (1991) identified principals as having the most supportive role, while McFerrin (1987) found special education teachers more supportive than regular education teachers in all areas of mainstreaming.

When both variables representing collaborative strategies were analyzed in this study, significant differences were found between perceptions of the regular education and the special education teachers. Special education teachers more strongly agreed with the "need for" and "importance of" collaborative strategies than the regular education teachers.

Respondents with 13 through 19 years experience most strongly agreed on importance of collaboration, consultation, and mutual planning time (the collaborative strategy subset). These respondents were at mid career. We expected more recent college graduates to most strongly agree since many have taken collaborative course work and many states now require a special education course for certification.

The analysis of position by school law courses yielded a statistically significant finding in the subset of perceived inclusion barriers. Principals perceived the conditions for inclusion as less prohibitive than the other two groups. Those responders with two or more courses in school law may have had more knowledge pertaining to barriers to inclusion. We expected this finding.

Number of years principals held administrative positions was statistically significant in the subset of factors supporting the integration of students with disabilities. Principals with the least years of experience (1-6 years) more strongly agreed with the supports for inclusion than did the other groups. This could have been a result of their more recent training and knowledge of school reform issues. McCaneny's (1992) findings were parallel, showing that more experienced principals were less inclined to mainstream students with disabilities.

Educators who had worked in the education field for 13-19 years more strongly agreed with the importance of collaborative strategies subset. Perhaps educators gain the confidence and insight to work with one another as they gain experience. Collaborative strategies means were higher than the means of the other subsets. Regular education teachers were the least in agreement with the collaborative strategies statements. Responses of regular education teachers may reflect the burden of trying to meet the needs of all students, particularly in light of the changing American classrooms. Principals may have a better over-all picture of schools; and special education teachers may have a clearer view of the abilities of students with disabilities. The importance of collaboration as a strategy for integration of students with disabilities was the highest ranked item in the survey.

Data from special education teachers yielded the highest means for inclusive education. Special education teachers may have had more exposure to the debate about inclusive education through their professional literature than the other two groups. Regular education teacher responses were the lowest in this category and were not as supportive of inclusive education as the other two groups. Principals and special education teachers were close in their response means. Rankings by position for this section were identical to Galis' findings for elementary school personnel in Georgia (1994).

The lowest means were found in the area of perceived barriers to inclusion. Data pertaining to principals reflected the lowest mean in this category. Regular education teachers had the highest mean response indicating that they perceived the choices provided as being greater obstacles to inclusive education. The top three perceived barriers were identified as lack of adequate amount of staff, lack of shared special/education planning time, and lack of amount of planning time allocated. These findings were similar to the barriers identified by Burello and Wright (1992) and needed competencies rated in a study by West and Cannon (1988). School climate, negotiations with teachers organizations, and school board policy had the lowest means, indicating that these factors presented the least inhibitions to inclusion. Funding issues were identified as major barriers by several researchers (Dempsey & Fuchs, 1993; McLaughlin & Owings, 1992; National Council on Disability, 1995), but respondents in this study did not perceive the weighted funding as a barrier nor flat funding as a support to inclusion.

The mean responses for perceived supports to inclusive education were clustered closely together. Special education teachers had a slightly higher mean response than the other two groups. The three supports with the highest mean scores were: funds for staff training, funds and/or release time for staff collaborative planning, and a lead teacher trained in special education and instructional strategies. These items were perceived to be the supports most helpful to an inclusive environment. The NCERI (1994) identified similar needs: staff training, collaborative support systems and time for such planning, along with visionary leadership. Wolery et al. (1995) identified the same priorities, labeling them training, meetings and support personnel.

All three items in the need for change (Section III) indicated strong agreement. "Training in modifications for students with disabilities who need adaptations in an instructional environment" was the highest ranked. The need for staff development for collaborative teaching and more opportunities for collaboration were also strongly supported. The response to these items appeared to indicate a willingness to develop skills to work with included students. Collaboration, and supports for staff and students were also determined to be necessary by the NCERI (Lipsky, 1994).

**Recommendations for Practice**

Respondents highly endorsed the importance of collaborative strategies. Total years in education was significant for respondents with 13-19 years experience. Perhaps those individuals could serve as mentors for their peers with less experience and encourage confidence in their abilities. Training in collaborative strategies and student modifications are strongly recommended.

Responders suggested that "Integration into general education classes is one of several strategies which should be considered for students with disabilities." This response, the highest ranked statement in Section II, indicates that they might have been weighing general education as one of the options for students with disabilities. Considering a continuum of services is also supported by case law and regulations.

The statement receiving the second highest agreement was: "It is important that behavioral expectations be maintained consistently for all students in a class, regardless of disability." Heumann (1994), Assistant Secretary for the Office of Special Education and Rehabilitative Services (OSERS), stated that one of the relevant factors to be used to determine if a placement was appropriate under IDEA was "the degree of disruption of the education of other students resulting in the inability to meet the unique needs of the student with a disability." (p. 3).Oberti v. Board of Education (1993) revealed that placement considerations could include an analysis of the possible negative effects of inclusion on other students in the class.

Students with disabilities should be provided the training and tools to manage their behavior. Models such as the one presented by Donaldson and Christiansen (1990) could provide suggestions for the development of a local school plan for assistance, behavior management, and instructional options for students with disabilities. Special education teachers should prepare students for reintegration in behavioral areas as well as academic areas. Programming for generalization to other environments must be included in that training. Monitoring for appropriate behaviors would be part of the ongoing assessment of students once re integrated.

Special education teachers could be used as a local school resource to provide training to the staff for appropriate behavioral strategies to be used. Students need concrete models of behavioral expectations for their successful behavioral integration into the regular classroom. Rock, Rosenberg, & Carran (1994) found that students with severe behavioral problems achieved higher reintegration rates when their former placement was in a program in a regular education school and zero to one mile(s) from the reintegration site.

The statement receiving the third strongest agreement was: "Students should be included in the general education environment to the greatest extent possible." This response appears to support inclusion even though practice does not currently reflect this at a high level for students in middle school settings. Perhaps models of inclusion should again be reviewed as in the case of the statement with which there was the strongest agreement.

The top three supports to inclusion were identified as funds for staff training, funds and/or release time for staff collaborative planning, and a lead teacher trained in special education and instructional strategies. The implementation of these strategies may serve to increase the inclusion of students and the success of individual students whose placement committee has identified the regular education classroom as the least restrictive environment. There are many proposals for staff development (Gallagher, 1994; Hamre-Nieptupski et al., 1990; Lipsky, 1994; National Council on Disability, 1995; Rath, 1989; Servatius, Fellows, & Kelly, 1989; Thousand & Villa, 1992; Villa, 1989). Training at the pre-service level in collaborative strategies might serve to provide new teachers with the skills for collaboration and the confidence that it can be implemented.

Conversely, the top three perceived barriers to inclusion were identified as lack of adequate amount of staff, lack of shared special/education planning time, and lack of amount of planning time allocated. Collaborative planning time was addressed in perceived supports. Middle schools historically have more planning time than other levels of education, so perhaps the issue may be more effective use of available planning time and time set aside specifically for collaborative teams. Parallel planning time can be established to address that concern. The issue of lack of staff was reported to be resolved when costs of transportation and more restrictive placements freed up funds for more personnel (National Council on Disability, 1995). Stainback and Stainback (1990) estimated that $20-$25 billion dollars were being spent annually on special education programs and that one in eight teachers in the U.S. was employed in special education. They asserted that these resources were adequate in terms of manpower and financial resources to provide support for facilitators to make inclusion work.

### Implications for Further Research

1. Findings from the review of literature for this study underscored the need for further efficacy studies of instruction for students with disabilities in a variety of settings, including both regular and special education classrooms. This was supported by the report of the National Council on Disability (1995) to the President. Christenson, Ysseldyke, & Thurlow (1989) reviewed the literature on critical instructional factors for students with mild disabilities and identified 10 instructional factors. Studies from sites incorporating those factors and promising practices identified by the National Council on Disability (1995) could possibly offer some answers to the efficacy questions.

2. The cost of educating a student with disabilities was approximated at 2.3 times that of a student without disabilities (Chaikind, Danielson,& Brauen,1993). Large amounts of federal, state, and local resources were spent on special education programs annually. Further study on the cost of inclusionary programs are needed since cost is often viewed as a barrier to such programs. Funding impacted the top three barriers identified in this study. Additionally, funding for students was often generated to the local school district based on the service delivery model, with no funding being provided for students with disabilities in the regular classroom (National Council on Disability, 1995). Financial incentives should be explored regarding inclusive settings.

3. Respondents with 13 through 19 years experience in education had significantly higher means than respondents in all other groups in the area of collaborative strategies. They more strongly supported collaborative strategies than individuals new to the profession. Galis (1994) identified educators with 17 to 21 years of experience as more positively supporting inclusive education. It would be beneficial for further study to explore the possible increased support for change by seasoned educators over persons in their first dozen years of the profession.

4. An analysis of possible middle school organizational patterns or structures that differentiate inclusion percentages from elementary school and high school settings would be beneficial. The U.S. Department of Education (1994) reported a dramatic decline in regular classroom settings for students with disabilities as they increased in age. The differentiation between elementary and middle schools remains a concern and analysis might reveal promising practices in the elementary school which could be successfully imported into the middle school setting.

5. Middle schools have historically integrated students with disabilities into non-academic classes often known as exploratories. Most students are successful in these non-academic classes, with the possible exception of students with emotional disorders (Rath, 1989). Further study of teaching strategies and management systems in these exploratory classes might be helpful to determine the supports given to students with disabilities in those settings that may not be provided in traditional academic classes. Some of the non-academic classes or exploratories did have academic components to them (such as foreign languages, computer, health, music theory, and art history).

### Concluding Statements

Respondents demonstrated support for the integration of students with disabilities into the regular education environment through their agreement with statements supporting inclusion as an effective strategy and a part of the continuum of services to be considered for LRE. There was support for collaborative strategies, provisions for staff training, and shared planning time. Behavioral expectations were identified as a concern when students with disabilities were included. The degree of disruption of the learning process for non disabled students has been viewed as an appropriate consideration in placement decisions in both the case law and by the Assistant Secretary for OSERS (Heumann, 1994). Cost considerations were not identified by the respondents as a priority among the possible perceived barriers, even though they were often cited as a concern in the literature. One school district reported that excess costs of inclusion were offset by savings in several areas, including transportation and fewer placements in out-of-district and more restrictive placements (National Council on Disability, 1995).

The literature review emphasized the principal as the pivotal change agent in school reform. Principals and special education teachers revealed statistically significant support for inclusion. Principal respondents reported a high level of input when planning took place for students with disabilities served in the regular classroom. Possible factors as barriers to inclusion were rated lower by principals in comparison to both regular and special education teachers when two or more courses in school law were taken. Rath (1989) identified three stages of integration of students with disabilities: inclusion, differentiation, and integration. The principal was viewed as the integrator since integration was a component of the larger organizational task of creating appropriate and effective integrative structures within the school.

We conducted this study to help answer the following question: What are the perceptions of front-line middle school educators regarding inclusion as a viable educational delivery system for students with disabilities? While we did not find a simple "yes" or "no" answer, indications are strong that there is a significant need to work with principals, teachers and special education teachers in middle schools if inclusion is to become fully accepted.

## References

Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research*, 64(2), 287-309.

Arrington, L. R. (1993). Integrating special education students into the regular classroom: An investigation and analysis of principal and teacher attitudes. Unpublished doctoral dissertation, The University of Arizona, Tucson.

BaiLey, D. & Winton, P. (11989). Friendship and acquaintance among families in a mainstreamed day care center. *Education and Training of the Mentally Retarded*; 24, 117-113.

Baines, L., Baines, C., & Masterson, C. (1994). Mainstreaming: One school's reality. *Phi Delta Kappan*, 76(1), 39-64.

Baker, E. T. (1994). Meta-analytic evidence for non-inclusive educational practices: Does educational research support current practice for special-needs students? Unpublished doctoral dissertation. Philadelphia: Temple University.

Baker, E. T., Wang, M. C., & Walberg, H. J. (1995). The effects of inclusion on learning.*Educational Leadership*,42(4), 33-35.

Bauwens, J., Hourcade, J. J., & Friend, M. (1989). Cooperative teaching: A model for general and special education integration.*Remedial and Special Education*, 10(2), 17-22.

Best, J. W. (1970).*Research in education*. Englewood Cliffs, NJ: Prentice-Hall.

Burrello, L. C. (Ed.). (1991). New role relationships between principals and special education administrators. Principal letters: Practices for inclusive schools. Bloomington, IN: Indiana University, National Academy/CASE.

Burrello, L. C., & Wright, P. T. (Eds.). (1992). Principal leadership. Principal letters: Practices for inclusive schools. Bloomington, IN: Indiana University, National Academy/CASE.

Byrnes, M. (1990). The regular education debate: A view from the field. *Exceptional Children*, 56(4), 345-349.

Carlberg, C. & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis.*Journal of Special Education*, 14, 295-309.

Chaikind, S., Danielson, L. C., & Brauen, M. L. (1993). What do we know about the costs of special education? A selected review. *The Journal of Special Education*, 26(4), 344-370.

Christenson, S. L., Ysseldyke, J. E., & Thurlow, M. L. (1989). Critical instructional factors for students with mild handicaps: An integrative review. *Remedial and Special Education*, 10(5), 21- 31.

Dempsey, S., & Fuchs, D. (1993). "Flat" versus "weighted" reimbursement formulas: A longitudinal analysis of statewide special education funding practices.*Exceptional Children*, 59(5), 433-443.

de Vaus, D. A. (1986).*Surveys in social research* (Contemporary Social Research Series, Vol. II). London: George Allen and Unwin Publishers Limited.

Donaldson, R., & Christiansen, J. (1990). Consultation and collaboration: A decision-making model.*Teaching Exceptional Children*, 22(2), 22- 25.

Farley, J. L. (1991). The current attitudes of principals and teachers regarding mainstreaming in Virginia middle-level schools. Unpublished doctoral dissertation, The College of William and Mary, Williamsburg, VA.

Fuchs, D., & Fuchs, L. (1994). Inclusive school movement and the radicalization of special education reform. *Exceptional Children*, 60(4), 294-309.

Fuchs, D., & Fuchs, L. (1995). What's "special" about special education? *Phi Delta Kappan*, 76(7), 522-530.

Galis, S. A. (1994). Inclusive education: Attitudes and beliefs of special and regular educators in the state of Georgia. Unpublished doctoral dissertation, University of Georgia, Athens.

Galis, S. A., & Tanner, C. K. (1995). Inclusion in elementary schools: A survey and policy analysis. *Education Policy Analysis Archives, [http://olam.ed.asu.edu/epaa/v3n15.html]* 3(15), 1-24.

Gallagher, J. J. (1994). The pull of societal forces on special education. *The Journal of Special Education*, 27(4), 521-530.

Gallup, G. (1976).*The sophisticated poll watcher's guide* (Rev. Ed.). Princeton: Princeton Opinion Press.

Giangreco, M. F., Dennis, R., Cloninger, C., Edelman, S., & Schattman, R. (1993). "I've counted Jon": Transformational experiences of teachers educating students with disabilities. *Exceptional Children*, 59(4), 359-373.

Green, A. & Stoneman, Z. (1989). Attitudes of mothers and fathers of nonhandicapped children. *Journal of Early Intervention*, 13(4), 292- 304.

Hasazi, S. B., Johnston, A. P., Liggett, A. M., & Schattman, R. A. (1994). A qualitative policy study of the least restrictive environment provision of the Individuals with Disabilities Education Act. *Exceptional Children*, 60(6), 491-507.

Heller, K., Holtzman, W., & Messick, S. (1982). Placing children in special education: A strategy for equity. Washington, D.C.: National Academy of Science Press.

Helmstetter, E., Peck, C., & Giangreco, M. F. (1993). Outcomes of interactions with peers with moderate or severe disabilities: A statewide survey of high school students. Unpublished manuscript, Washington State University.

Heumann, J. E. (1994, September 16). Answers to frequently asked questions about the requirement of The Individuals With Disabilities Education Act (IDEA). Available from U. S. Dept. of Education, OSERS, 400 Maryland Ave., S. W., Washington, D.C. 20202.

Hobbs, N. (1980). An ecologically oriented service-based system for the classification of handicapped children. In E. Salzinger, J. Antrobus, & Glick (Eds.), *The Ecosystem of the "risk" child* (pp. 271-290). New York: Academic Press.

Huefner, D. S. (1994). The mainstreaming cases: Tensions and trends for school administrators. *Educational Administration Quarterly*, 30(1), 27-55.

Individuals with Disabilities Education Act 20 U.S.C. $ 1400, et seq.

Jenkins, J. R., Jewell, M., Leicester, N., OUConnor, R. O., Jenkins, L. M., & Troutner (1994). Accommodations for individual differences without classroom ability groups: An experiment in school restructuring.*Exceptional Children*, 60(4), 344-358.

Johnston, D., Proctor, W., & Corey, S. (1995). Not a way out: A way in. *Educational Leadership*, 42(4), 46-49.

Jones, M. M., & Carlier, L. L. (1995). Creating inclusionary opportunities for learners with multiple disabilities: A team-teaching approach. *Teaching Exceptional Children*, 27 (3), 23-27.

Lipsky, D. K. (1994). National survey gives insight into inclusive movement.*Inclusive Education Programs*, 1 (3), 4-7.

MacKinnon, J. D., & Brown, M. E. (1994). Inclusion in secondary schools: An analysis of school structure based on teacher' images of change. *Educational Administration Quarterly*, 30(2), 126-152.

Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal*, 30(1), 123-148.

Maroldo, R. A. (Ed.). (1994). Heterogeneity: Strategies for accepting differences on the general classroom. *Inclusive Education Programs*, 1(4), 8-10.

McCaneny, F. X. (1992). The impact of school principals' attitudes toward mainstreaming on student referral. Unpublished doctoral dissertation, Temple University, Indiana.

McDonnell, A. P., & Hardman, M. L. (1989). The desegregation of America's special schools: Strategies for change. *Journal of the Association for Persons with Severe Handicaps*, 14(1), 68-74.

McFerrin, A. J. (1987). A comparison of the attitudes of regular and special education teacher education faculty toward mainstreaming. Unpublished doctoral dissertation, University of Georgia, Athens. Meyen, E. (1978). An introductory perspective. In E. Meyen (Ed.), *Exceptional children and youth* (pp. 2-64). Denver: Love Publishing.

Miller, L. (1990). The regular education initiative and school reform: Lessons from the mainstream. *Remedial and Special Education*, 11(3), 17-22.

Morse, W. C. (1994). Comments from a biased viewpoint. *The Journal of Special Education*, 27(4), 531-542.

National Center on Educational Restructuring and Inclusion (1994). National survey on inclusive education. Bulletin: NCERI, 1, 1-4.

National Commission on Excellence in Education (1983). A nation at risk. Arlington, VA.: ERIC Document Reproduction Service.

National Council on Disability (1995). Improving the implementation of the Individuals with Disabilities Education Act: Making schools work for all of America's children. Washington, D.C.: Author.

Nevin, A., Thousand, J., Paolucci-Whitcomb, P., & Villa, R. (1990). Collaborative consultation: Empowering public school personnel to provide heterogeneous schooling for all -- or, who rang that bell? *Journal of Educational and Psychological Consultation*, 1(1), 41-67.

Nunnery, M. Y., & Kimbrough, R. B. (1971). *Powers, polls, and school elections*. Berkeley, CA: McCutchan.

Oberti v. Board of Education of the Borough of Clementon School District, 789 F. Supp. 1322, 75 Ed.Law Rep. 258, (D.N.J. 1992); 801 F. Supp. 1393 (D.N.J. 1992); affirmed 995 F.2d 1204, 83 Ed.Law Rep. 1009 (3d Cir. 1993).

O'Neil, J. (1995). Can inclusion work? A conversation with Jim Kauffman and Mara Sapon-Shevin. *Educational Leadership*, 42(4), 7-11.

Osborne, A. G., & Dimattia, P. (1994). The IDEA's least restrictive environment mandate: Legal implications. *Exceptional Children*, 61(1), 6-14.

Peck, C. A., Carlson, P., & Helmstetter, E. (1992). Parent and teacher perceptions of outcomes for typically developing children enrolled in integrated early childhood programs: A statewide survey. *Journal of Early Intervention*, 16, 53-63.

Peck, C. A., Mabry, L., Curley, J., & Conn-Powers, M. (May 1993). Implementing inclusion at the early childhood level: A follow-along study of 54 programs. Paper presented at the annual Washington State Infant and Early Childhood Conference, Seattle.

Principals favor reconsideration of inclusion policy for disabled (1995). Report on Education of the Disadvantaged, Sample Issue, p. 2.

Raison, J., Hanson, L. A., Hall, C., & Reynolds, M. C. (1995). Another school's reality. *Phi Delta Kappan*, 76(6), 480-482.

Rath, C. C. (1989). The integration of handicapped students in the middle school: The principal's view. Unpublished doctoral dissertation, Boston University.

Reynolds, M., & Birch, J. (1982). Teaching exceptional children in all America's schools (2nd ed.). Reston, VA: The Council for Exceptional Children.

Riley, R. W. (1993, September). A new direction for education. *Principal*, 73, 5-8.

Rock, E. E., Rosenberg, M. S., & Carran, D. T. (1994). Variables affecting the reintegration rate of students with serious emotional disturbance. *Exceptional Children*, 61(3), 254-268.

Sapon-Shevin, M. (1990). Student support through cooperative learning. In W. Stainback & S. Stainback (Eds.), Support networks for inclusive schooling: Interdependent integrated education. Baltimore, MD: Paul H. Brookes.

Schiller, E. P., Coutinho, M., & Kaufman, M. J. (1993). Establishing a unity between regular and special education. In Integrating Students with Special Needs. Washington, D.C.: National Education Association.

Schulte, A. L., Osborne, S. S., & McKinney, J. D. (1990). Academic outcomes for students with learning disabilities in consultation and resource programs. *Exceptional Children*, 57(2), 162-172.

Schumm, J. S., & Vaughn, S. (1992). Planning for mainstreamed special education students: Perceptions of general education teachers. *Exceptionality*, 3(2), 81-98.

Servatius, J. D., Fellows, M., & Kelly, D. (1989). Schools are for all kids: The leadership challenge. Unpublished training manual, San Francisco State University, CRI, San Francisco, CA.

Shapiro, J. P., Loeb, P., Bowermaster, D., Wright, A., Hadden, S., & Toach, T. (1993, December 13). Separate and unequal. *U. S. News and World Report*, p. 46-60.

Slavin, R. E., Madden, N. A., Karweit, N. L., Dolan, L., Wasik, B. A., Shaw, A., Mainzer, K. L., & Haxby, B. (1991). Neverstreaming: Prevention and early intervention as an alternative to special education. *Journal of Learning Disabilities*, 24(6), 373-378.

Slavin, R.E., Madden, N. A., Karweit, N. L., Livermon, B. J., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal*, 27(2), 255-278.

Smith v. Robinson, 468 U.S. 992, 1009 (1984).

Stainback, S., & Stainback, W. (1990). Inclusive schooling. In W. Stainback & S. Stainback (Eds.), *Support networks for inclusive schooling: Independent integrated education*. Baltimore, MD: Paul H. Brookes.

Stainback, S., & Stainback, W. (1984). A rationale for the merger of special and regular education. *Exceptional Children*. 51(2), 102- 111.

Staub, D., & Peck, C. A. (1995). What are the outcomes for non disabled students? *Educational Leadership*, 52(4), 36-41.

Thousand, J. S., & Villa, R. A. (1992). Collaborative teams: A powerful tool in school restructuring. In R. A. Villa, J. S. Thousand, W. Stainback, & S. Stainback (Eds.),*Restructuring for caring and effective education: An administrative guide to creating heterogeneous schools*. Baltimore, MD: Paul H. Brookes.

U. S. Department of Education (1994). To assure the free appropriate education of all children with disabilities: 16th Annual Report to Congress on the implementation of IDEA. Washington, D.C.: U.S. Department of Education.

Van Dyke, R., Stallings, M. A., & Colley, K. (1995). How to build an inclusive school community: A success story. *Phi Delta Kappan*, 76(6), 475-479.

Vergason, G. A., & Anderegg, M. L. (1992). Preserving the least restrictive environment. In W. Stainback & S. Stainback (Eds.),*Controversial issues confronting special education*.

Needham Heights, MA: Allyn and Bacon.

Villa, R. A. (1989). Model public school inservice programs: do they exist? *Teacher Education and Special Education*, 12, 173-176.

Villa, R. A., & Thousand, J. S. (1992). Student collaboration: an essential for curriculum delivery in the 21st Century. In S. Stainback & W. Stainback (Eds.), *Curriculum considerations in inclusive classrooms: Facilitating learning for all students*. Baltimore, MD: Paul H. Brookes.

Wade, B., & Moore, M. (1992). *Patterns of educational integration: International perspectives on mainstreaming children with special educational needs.* Wallingford, Oxfordshire UK: Triangle Books.

Walker, L. J. (1987). Procedural rights in the wrong system: Special education is not enough. In A. Gartner and T. Joe (Eds.). *Images of the Disabled / Disabling Images*. New York: Praeger.

Wang, M. C., & Baker, E. T. (1985-86). Mainstreaming programs: Design features and effects. *The Journal of Special Education*, 19, 503- 521.

West, J. F. & Cannon, G. S. (1988). Essential collaborative consultation competencies for regular and special educators. *Journal of Learning Disabilities*, 21(1), 57-63.

White, L. L. (1993). Cobb County Schools principals' perception of inclusion. Unpublished education specialist degree project, West Georgia College, Carrollton, GA.

Yatvin, J. (1995). Flawed assumptions. *Phi Delta Kappan*, 76(6), 482-484.

Ysseldyke, J. E., & Algozzine, B. (1982). *Critical issues in special and remedial education.* Boston: Houghton Mifflin.

## Appendix

### Table 3
### Individual Items by Mean
### for All Respondents by Cluster (Section II)

| Statement | Item/ Variable | *N | Mean | Standard Deviation |
|---|---|---|---|---|
| Inclusive Education:<br>(6 point Likert scale: 1=Strongly Disagree to 6=Strongly Agree) | | | | |
| Integration is generally an effective strategy for mild disabilities | 1/16 | 704 | 4.869 | 1.105 |
| I have input into programming for students with disabilities | 2/17 | 703 | 4.599 | 1.455 |
| Maximum class size should be reduced | | | | |

| | | | | |
|---|---|---|---|---|
| when including students with disabilities | 3/18 | 703 | 5.131 | 1.151 |
| Integration can be beneficial to other students | 4/19 | 704 | 4.700 | 1.164 |
| Students should be served in reg. ed. regardless of disability | 5/20 | 703 | 2.996 | 1.570 |
| Opportunities to plan on a regular basis with colleagues | 6/21 | 704 | 3.712 | 1.650 |
| Behavioral expectations should be the same for all students | 7/22 | 705 | 4.955 | 1.329 |
| Reg. ed. teachers must devote most of their time with included students | 8/23 | 703 | 3.073 | 1.445 |
| Students should be included to the maximum extent possible | 9/24 | 704 | 4.984 | 1.207 |
| Integration will limit progress of students with disabilities | 10/25 | 701 | 2.688 | 1.408 |
| Students with disabilities are disruptive to reg. ed. classes | 11/26 | 702 | 2.984 | 1.400 |
| Integration is one of several strategies to consider | 12/27 | 702 | 5.057 | 1.119 |
| Collaborative Strategies: | | | | |
| (6 point Likert scale: 1=Little to 6=Extensive) | | | | |
| Support for change: more time for collaboration | 13/28 | 702 | 4.802 | 1.143 |
| Support for change: staff development about collaboration | 14/29 | 702 | 4.960 | 1.146 |
| Support for change: training in modifications for included students | 15/30 | 702 | 5.081 | 1.048 |
| (6 point Likert scale: 1=not important to 6=very important) | | | | |
| Importance to integration: | | | | |

| | | | | |
|---|---|---|---|---|
| collaboration | 39/54 | 701 | 5.515 | .827 |
| Importance to integration: co-teaching | 40/55 | 702 | 4.950 | 1.236 |
| Importance to integration: consultation | 41/56 | 700 | 5.114 | 1.106 |
| +Importance to integration: reduced class size | 42/57 | 703 | 5.134 | 1.167 |
| Importance to integration: mutual planning time | 43/58 | 703 | 5.448 | .891 |
| Factors perceived as barriers: | (6 point Likert scale: 1=Least Inhibiting to 6=Most Inhibiting) | | | |
| Concern for student outcomes | 16/31 | 694 | 3.280 | 1.464 |
| Role responsibility gen./reg. ed. teacher | 17/32 | 699 | 3.794 | 1.432 |
| Federal rules/regs | 18/33 | 693 | 3.420 | 1.647 |
| Lack of staff | 19/34 | 696 | 4.503 | 1.512 |
| Lack of central office support | 20/35 | 693 | 2.846 | 1.623 |
| Lack of staff willingness | 21/36 | 694 | 3.605 | 1.546 |
| Teacher unions | 22/37 | 673 | 2.095 | 1.390 |
| Planning time constraints (time) | 23/38 | 690 | 4.291 | 1.568 |
| Planning time not shared | 24/39 | 694 | 4.419 | 1.617 |
| School board policies | 25/40 | 682 | 1.999 | 1.325 |
| School climate | 26/41 | 693 | 2.473 | 1.447 |
| State rules & regs | 27/42 | 685 | 2.761 | 1.607 |
| weighted funding | 28/43 | 660 | 2.986 | 1.640 |
| Factors indicating perceived support: | (6 point Likert scale: 1=Least Helpful to 6=Most Helpful) | | | |
| Asst. principal as a generalist | 30/45 | 687 | 3.154 | 1.802 |
| Central office support | 31/46 | 694 | 3.903 | 1.725 |
| De-emphasis on test | | | | |

| | | | | |
|---|---|---|---|---|
| scores (standardized) | 32/47 | 695 | 4.187 | 1.502 |
| Flat funding | 33/48 | 676 | 3.621 | 1.792 |
| Funding/release time for collaborative training | 34/49 | 689 | 5.250 | 1.070 |
| Funds for staff training | 35/50 | 699 | 5.280 | .976 |
| Lead teacher trained in spec. ed. & instruction | 36/51 | 695 | 4.994 | 1.174 |
| School board support | 37/52 | 689 | 4.224 | 1.679 |

* The number of respondents varies because of missing cases.
+ Item was dropped from the subset based on the reliability analysis.

## About the Authors

### C. Kenneth Tanner

Professor
Department of Educational Leadership
The University of Georgia
Athens, GA 30603

Kenneth Tanner is a professor in the Department of Educational Leadership at the University of Georgia, Athens, GA. He earned an Ed. D. from the Florida State University in educational administration and business management and holds membership in AERA, ISEP, and CEFPI. He has been recognized as a Danforth-Johnson Scholar by Stanford University and has published 4 books and 60 articles. His areas of research and teaching are educational policy analysis, school environmental desin and planning. He may be reached at any of the following: ktanner@moe.coe.uga.edu; or Department of Educational Leadership, UGA, Athens, GA 30602; or FAX (706) 542-5873, Phone (706) 542-4067.

### Deborah Jan Vaughn Linscott

Instructional Support Teacher
Fulton County Schools
2816 Briarwood Boulevard
East Point, GA 30344

Deborah Jan Vaughn Linscott is a special education teacher in Fulton County, Georgia Public Schools. She received the Ed. D. In Educational Leadership from the University of Georgia and the B. A. From the University of Arizona. Her areas of interest include the effects of inclusive education on students and methods of teaching special education. She may be reached at (770) 445-5031.

### Susan Allan Galis

Special Education Director
Commerce City Schools
Commerce, GA 30529

Susan Allan Galis received her Ed. D. From the University of Georgia in Special Education. Currently she serves as Special Education Director for Commerce City Jefferson City, GA Public Schools. In addition she is a part-time Assistant Professor in the Department of Special Education at the University of Georgia. Dr. Galis received her BM degree in music therapy from Florida State University and is a registered music therapist. She is particularly interested in special education policy issues. She may reached at (706) 335-5500 or (706) 367-2883

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

*Education Policy Analysis Archives* are "gophered" at olam.ed.asu.edu.

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

## Editorial Board

Greg Camilli
*camilli@rci.rutgers.edu*
Andrew Coulson
*a_coulson@msn.com*
Sherman Dorn
*dorn@typhoon.coedu.usf.edu*
Thomas F. Green
*tfgreen@mailbox.syr.edu*
Arlen Gullickson
*gullickson@gw.wmich.edu*
Aimee Howley
*ess016@marshall.wvnet.edu*
William Hunter
*hunter@acs.ucalgary.ca*
Benjamin Levin
*levin@ccu.umanitoba.ca*
Dewayne Matthews
*dm@wiche.edu*
Les McLean
*lmclean@oise.on.ca*
Anne L. Pemberton
*apembert@pen.k12.va.us*
Richard C. Richardson
*richard.richardson@asu.edu*
Dennis Sayers
*dmsayers@ucdavis.edu*
Robert Stonehill
*rstonehi@inet.ed.gov*

John Covaleskie
*jcovales@nmu.edu*
Alan Davis
*adavis@castle.cudenver.edu*
Mark E. Fetler
*mfetler@ctc.ca.gov*
Alison I. Griffith
*agriffith@edu.yorku.ca*
Ernest R. House
*ernie.house@colorado.edu*
Craig B. Howley
*u56e3@wvnvm.bitnet*
Richard M. Jaeger
*rmjaeger@iris.uncg.edu*
Thomas Mauhs-Pugh
*mauhsput@rocky.edu*
Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*
Susan Bobbitt Nolen
*sunolen@u.washington.edu*
Hugh G. Petrie
*hgpetrie@acsu.buffalo.edu*
Anthony G. Rud Jr.
*rud@purdue.edu*
Jay Scribner
*jayscrib@tenet.edu*
Robert T. Stout
*aorxs@asuvm.inre.asu.edu*

**Education Policy Analysis Archives**

**Volume 4 Number 20** **December 24, 1996** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.

Editor: Gene V Glass,Glass@ASU.EDU. College of Education, Arizona State University,Tempe AZ 85287-2411

Copyright 1996, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

# *The Bell Curve*: Corrected for Skew

**Haggai Kupermintz**
**Stanford University**

**Abstract** This commentary documents serious pitfalls in the statistical analyses and the interpretation of empirical evidence presented in *The Bell Curve*. Most importantly, the role of education is re-evaluated and it is shown how, by neglecting it, *The Bell Curve* grossly overstates the case for IQ as a dominant determinant of social success. The commentary calls attention to important features of logistic regression coefficients, discusses sampling and measurement uncertainties of estimates based on observational sample data, and points to substantial limitations in interpreting regression coefficients of correlated variables.

## Introduction

*The Bell Curve* by Richard Herrnstein and Charles Murray (henceforth H&M) puts forward a strong thesis about the centrality of intelligence in determining contemporary American social structure. Following its publication in October 1994, *The Bell Curve* sparked an intense public debate over its assertions, methodology and conclusions. Most of the book's critics, in a flood of newspaper articles, TV talk shows, academic journal articles and a few books, focused on *The Bell Curve*'s treatment of ethnic and racial group differences in intelligence, the role of heredity in determining these differences, and the social and political agenda advocated by H&M. The heated debate was clearly another wave of the controversies about genes, IQ and public policy (see, e.g., Cronbach, 1975).

*The Bell Curve* is distinguished by its extensive use of statistical analyses to support a strong social theory. Other authors have provided critical examination of some statistical and measurement aspects of *The Bell Curve*, raising concerns about the appropriateness of causal inferences, model specification (most notably the absence of measures of education from the models), model fit and the validity of IQ and SES measures, among other issues. Some of these concerns will be echoed here in detail. The current commentary will go beyond delineation of these issues in principle or theory, to reexamine the statistical evidence and to analyze further the data presented in *The Bell Curve*.

H&M explore the relationship between social stratification and the distribution of cognitive abilities which, according to their thesis, will inevitably lead to a "world in which cognitive ability is the decisive dividing force" (p.25). Part I of the book is devoted to an elaborate exposition of the emergence and the increasing isolation of a "cognitive elite", driven by radical transformations in educational, occupational and economic forces in American society throughout the twentieth century. What are the consequences of this current American landscape that has been stratified so forcefully according to cognitive ability?

In part II of the book, H&M launch a series of statistical analyses to examine the role of intelligence, as measured by an IQ test, in determining a myriad of social ailments such as poverty, school dropout, unemployment and labor force dropout, welfare dependency and criminal behavior. The analyses of part II use a sub-sample of non-Latino white respondents from the National Longitudinal Survey of Youth (NLSY)--a nationally representative sample of 12,686 young men and young women who were 14 to 22 years of age when they were first surveyed in 1979. By focusing on the white sub-sample, H&M argue that "cognitive ability affects social behavior without regard to race and ethnicity" (p. 125). Only later, in Part III, when the importance of intelligence as a powerful determinant of social behavior has been allegedly demonstrated, do H&M turn to examine ethnic and racial group differences. An evaluation of the scientific merit of the book will best be served by focusing on how H&M handle and present the less controversial evidence about the role of intelligence in the lives of young white Americans. As Charles Murray notes, "perhaps the most important section of *The Bell Curve* is Part II" (1995, p. 27). Indeed, many of the arguments and conclusions to appear later in the book rely heavily on the success of the case made in Part II, which constitutes (together with Appendices 2,3,4) a dense collection of statistics, tables, graphs, and technical details. H&M use the case of poverty, presented in Chapter 6, to "set the stage for the social behaviors to follow" (p.125). This chapter provides a basic template for their formulation of research questions, analysis strategies and use and interpretation of statistical methods. As such, it will be appropriate to focus here in some detail on this chapter. Chapter 6 asks, "What causes poverty?", or more specifically, "If you have to choose, is it better to be born smart or rich?" (p.127). Let us examine how H&M arrive at what they claim is an "unequivocal" answer: "smart".

## Logistic Regression Coefficients

The basic analytical tool H&M employ is a set of multiple regression equations. The independent variables are IQ, SES, and age. (Age is included in the models because of the nature of the NLSY sample. It is inconsequential to the arguments presented here and will not be further discussed.) The IQ test used throughout *The Bell Curve* is the Armed Forces Qualification Test (AFQT), a subset of the Armed Services Vocational Aptitude Battery (ASVAB). The SES measure is an average of standardized parental education, parental occupation, and family income. The dependent variable is whether a respondent in the NLSY was below the poverty line in 1989. H&M examine the regression results: they observe that the IQ regression coefficient (-.84) is much larger than the SES coefficient (-.33); they then plot a graph showing how the probability of being in poverty is predicted by the model as a function of IQ or SES, holding the other variable constant at its average value. (The regression equation is given in p. 596, and the graph in p. 134.) H&M conclude: "Cognitive ability is more important than parental SES in determining poverty" (p.135), independent of any role SES might play in determining the likelihood of poverty. How warranted is this conclusion?

For those not versed in the details of regression analysis, H&M provide a primer in Appendix 1 (pp. 553-577) entitled: "Statistics for People Who Are Sure They Can't Learn Statistics." After explaining basic statistical concepts, multiple linear regression is introduced. Logistic regression, the technique employed throughout Part II, is presented as a simple adaptation of linear regression to handle binary outcomes: "It tells us how much change there is in the probability of being unemployed, married, and so forth, given a unit change in any given variable, holding all the other variables in the analysis constant" (p. 567). The unsuspecting reader misses one important point: The value chosen at which to "hold a variable in the analysis constant" has a direct impact on the magnitude of anticipated change in the probability of the outcome, given a unit change in any other variable. H&M identify the mathematical function responsible for this behavior of the logistic regression, the log odds, or logistic function, later in the introduction to the results in Appendix 4, but they are silent about its consequences. As we shall see, this seemingly insignificant technical point has crucial implications for the interpretation of logistic regression results on a probability scale.

Let us examine what happens when we use the same regression coefficients, the same model, but decide to hold SES at other values than its average. Should we expect to see any noticeable difference in the relations between IQ and the probability of being in poverty? After all, we are still holding SES constant, and, as H&M assure us, "here is the relationship of IQ to social behavior X after the effects of socioeconomic background have been extracted" (p. 123).

Figure 1 depicts the predicted probabilities of being in poverty as a function of IQ at three values of SES: the SES average (the one shown in *The Bell Curve*), and 2 standard deviations above and below the SES average. Contrary to what we might have expected after being told that the effects of SES has been extracted out, the effect of IQ on the probability of being in poverty is much stronger when SES level is lower; it is much weaker when SES level is higher! This is a necessary consequence of the nature of the logistic regression model. For persons with lower socioeconomic status, the anticipated change in the likelihood of being poor associated with a unit change in IQ, is much larger than for those with higher socioeconomic status. This means that the risk of poverty induced by having lower intelligence is far more pronounced under conditions of adverse family environment. On the other hand, the privileges of a sound family background seem to mitigate the harsh consequences of lacking in cognitive abilities.

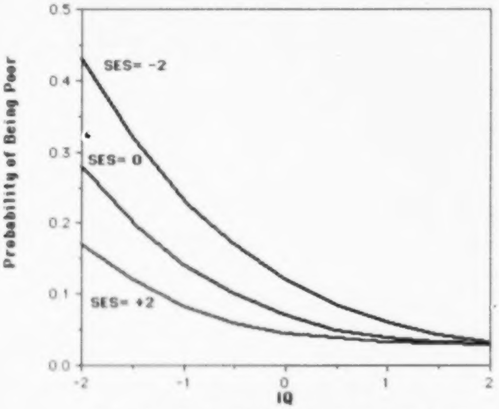

Figure 1. Probability of Being in Poverty as as a Function of Three SES Levels

Take for example two persons, a "smart" with an IQ of 115 (one standard deviation above the average), and a "dull" with an IQ of 85 (one standard deviation below the average). How do they compare in their respective risks of being poor? If they both come from an extremely poor background, the "dull" person is 18% more likely to be in poverty than the "smart"; On the other hand, if they both come from a family of extremely high socioeconomic status, the difference shrinks to only 6%. If we return to H&M original assertion about the logistic regression coefficient as indicating how much change will occur in the probability of poverty, given a unit change in IQ, we find that a two-units change (moving from -1 to 1 in standard deviations) in IQ, means three times more change in the probability of being poor for those with low SES compared with those with high SES. So much for "holding all the other variables in the analysis constant".

Clearly, Figure 1 tells a more complicated story than the one H&M would have the student of their statistics primer believe on the basis of interpreting the logistic regression coefficients as if they were linear or additive. Even more experienced researchers, who routinely run linear regression analyses, need more than what H&M are willing to provide as a guide to the proper interpretation of their logistic regression results. In the authoritative source on Generalized Linear Models, of which logistic regression is a special case, McCullagh and Nelder (1989) provide such guidance, as well as call attention to the fact that "...statements given on the probability scale are more complicated because the effect on [the probability of an outcome] of a unit change in X2 *depends on the values of X1 and X2*" (p. 110; italics added). In discussing the "special case of education" (we shall have more to say on this later), H&M quite rightly assert that "...to take education's regression coefficient seriously tacitly assumes that intelligence and education could vary independently and produce similar results.

No one can believe this to be true in general: indisputably, giving nineteen years of education to a person with IQ of 75 is not going to have the same impact on life as it would for a person with an IQ of 125" (p.125). Why should we, then, take the IQ regression coefficient seriously when, as we just saw, having a high (or low) IQ for a person coming from a poor background is not going to have the same impact on life as for a person coming from a wealthy background?

Let us now review the substantive conclusion H&M draw from the regression results: "If a white child of the next generation is given a choice between being disadvantaged in socioeconomic status or disadvantaged in intelligence, there is no question about the right choice" (p. 135). Indeed, there is no question: If your parents are rich enough, you can afford to be very dull and still can expect to escape poverty. If, on the other hand, you made the poor (literally) choice of being born to a low SES family, chances are that intellectual weakness will carry grave consequences for you. This, of course, is a caricature of serious hypothesizing about the dynamics of cognitive abilities and social conditions, but it brings us to the next issue--the independence (or the lack thereof) of independent variables.

### Independence of Independent Variables

H&M point out that "variables that are closely related can in some circumstances produce a technical problem known as *multicollinearity*, whereby the solutions produced by regression equations are unstable and often misleading" (pp.124-125; italics in original). Attention to potential effects of *multicollinearity* (meaning simply that the independent variables are correlated with each other), is indeed warranted when dealing with an attempt to disentangle via statistical analysis the effects of variables that are highly correlated in nature. Observing correlations of .50 and .64 between education and SES and IQ, respectively, cause H&M to raise a concern about the interpretation of a regression model that includes all three of them as independent variables. But what about the association between SES and IQ? Are they free to vary independently? Are they sufficiently uncorrelated as not to sound a similar alarm?

The correlation between the AFQT scores and parental SES in the NLSY data is .55. After reporting this correlation, H&M summarize: "Being brought up in a conspicuously high-status or low-status family from birth probably has a significant effect on IQ, independent of the genetic endowment of the parent" (p. 589). Although the magnitude of these effects or their explanation are debatable, the IQ scores used in *The Bell Curve* to demonstrate the independent role of a cognitive endowment are caused to an important degree by parent's SES. This means, to rephrase H&M argument about ignoring years of education in their regressions, that when IQ is used as an independent variable, it is to some extent expressing the effects of SES in another form. Can this be solved by the machinery of multiple regression? It is too often believed that regression analysis provides the proper statistical control, "accounting for" is the usual term, which mathematically remedies the confounding of effects imposed by the realities of the investigated phenomenon or by the study design. The answer is an unequivocal "No." Neter, Wasserman, and Kutner (1990) explain:

> "Sometimes the standardized regression coefficients, b1 and b2, are interpreted as showing that X1 has a greater impact on the [outcome variable] than X2 because b1 is much larger than b2. However, ...one must be cautious about interpreting regression coefficients, whether standardized or not. The reason is that when the independent variables are correlated among themselves, as here, the regression coefficients are affected by the other independent variables in the model." (By a happy circumstance, the correlation alluded to in this section is 69, almost exactly the correlation between IQ and SES!) "Hence, it is ordinarily not wise to interpret the magnitudes of standardized regression coefficients as reflecting the comparative importance of the independent variables" (p.294).

For a detailed discussion of these issues, the reader is invited to consult Chapter 13 of Mosteller & Tukey's *Data Analysis and Regression* (1977). They masterfully demonstrate the problems of interpreting regression coefficients, and sound very clear warnings concerning the comparison of regression coefficients even for fully deterministic systems under tight experimental control.

### A Scale is a Scale is a Scale?

The correlation between independent variables is not the only factor affecting the magnitude, and consequently the interpretation, of linear or logistic regression coefficients. It is important to recognize the effects on estimated regression parameters due to errors of measurement. H&M go into great detail to document the superior measurement qualities of their IQ test - the AFQT. That the AFQT provides good measurement of g, general cognitive ability, is demonstrated by high correlations among its four constituent tests, by high correlations with other measures of general ability, and by high loadings on the general factor of the ASVAB battery. (The latter is purported to represent g in common psychometric practice. It is interesting to note, however, that Gustafsson and Muthen (1994) show that the ASVAB lacks measures of Fluid Intelligence and its general factor is closer to Crystallized Intelligence, which they interpret as a broad verbal factor, closely associated with academic achievement.) The conclusion is that the AFQT is an exceptionally high quality instrument.

What, then, are the measurement qualities of the measure of socioeconomic status? Compared with the treatment of the AFQT scale, only meager information is presented to allow evaluating the quality of the SES scale. However, from the two pieces of information that are presented, a reliability coefficient of .76 and correlations among the four indicators comprising the scale ranging from .36 to .63, we can safely conclude that the SES measure is substantially inferior as a measurement device and is subject to considerable error. Moreover, for more than a quarter of the subjects only three of the indicators were available, further compromising the reliability of the scale. Therefore, "one must conclude that as a proxy for 15 years of environment, this is a variable measured with substantial error" (Delvin et al., 1995, p. 1468). The effect of the SES scale's low reliability on the regression results is quite clear: an underestimation of the SES effect run in a "horse race" against IQ. It is likely that the real differences between the effects of SES and IQ on the poverty in the population are smaller than what is reflected in H&M's estimates. In addition to errors of measurement, statistical uncertainties related to sampling are another major source of caution.

### Uncertainty in Statistical Estimates

Based on the logistic regression results, as depicted by the plots they draw, H&M make two strong predictions to demonstrate the different roles IQ and SES play in determining poverty. Paying attention to the far left-hand side of the plots on p. 134, we can observe that a white person from an unusually deprived socioeconomic background, with an average IQ, has a probability of about 11% of being in poverty. On the other hand, an extremely dull person with an average SES, has a probability of about 26% of being in poverty - more than double. Notice that these prediction use extreme values of IQ and SES to produce dramatic differences.

How accurate are these statements? How much confidence should we have that the real proportions in the population are close to the ones suggested by the statistical model estimated for this particular sample? An appropriate indicator of statistical uncertainty is the confidence interval of prediction. It informs us about the range of likely values we expect to encounter if we were to sample again from the same population. Confidence intervals for prediction in logistic regression models are easily obtained by using conventional methods (see Agresti, 1990, Chapter 12) or alternatively, by utilizing a computer intensive resampling technique known as bootstrapping (see Efron & Tibshirani, 1993).

Using both methods, we may compute confidence intervals for the two predictions above (at the 95% confidence level). The range of plausible values for a person from a deprived socioeconomic background with an average IQ goes from 8% to 16%. The range of plausible values for a dull person with average SES goes from 20% to 35%. (Both methods gave similar results.) The confidence interval for the difference between the two predictions indicates that this difference can be as small as 6% or as big as 26%.

Evidently, *The Bell Curve* ascribes unwarranted precision to estimates that are subject to considerable sampling error. The dramatic difference between the two estimates becomes much less so when one takes into account the statistical uncertainty associated with them. Thus when H&M declare categorically that the odds of poverty for a person with low IQ and average SES are "more than twice as great as the odds facing the person from deprived home but with average intelligence" (p.135), one needs to exercise great caution before accepting it on face value. But then, H&M themselves acknowledge (though only in a footnote) the complexities involved in comparing the magnitude of effects in multiple regression and promise: "We refrain from precise numerical estimates of how much more important IQ is than socioeconomic background..." (note 13, p.691).

We may also ponder: How valid is a comparison between a person with an IQ score of about 70 (two standard deviations below the average) and a person from a very poor family? That people with very low cognitive capacity face severe limitations in life is hardly a surprising or a fresh finding. For example, Jensen states that "most persons with any experience in the matter would agree that those with IQs below 70 or 74 have unusual difficulty in school and in the world of work. Few jobs in a modern industrial society can be entrusted to persons below IQ 70 without making special allowances for their mental disability" (1981, p.12). We should also remember that the youth falling into what H&M call Cognitive Class V, the very dull, are also routinely afflicted by severe socioeconomic conditions--they are on average almost an entire standard deviation below the mean in SES. The very dull are also the very poor. Attempts to disentangle the independent effects of cognitive ability and harsh environment are doomed, not because of technical complications, but because American social reality is less than generous towards its weakest citizens. It seems that *The Bell Curve* has no new story to tell here, but presenting such an extreme situation as an example of the general effect of IQ on social consequences is neither informative nor especially valid.

### The Special Case of Education

The impact of omission of important variables from a regression equation is widely recognized. Not only do the effects of the omitted variables cannot estimated, but other effects in the models might be biased and misinterpreted when an included independent variable is meaningfully correlated with an omitted one. Therefore, the absence of a measure of educational attainment from regression models set out to explain the likelihood of poverty, unemployment, welfare dependency and the likes, seems immediately curious. After all, education is the primary social institution responsible for providing the basic skills needed for a productive civil participation. The NLSY contains data on years of education respondents completed by 1990, which seems to be a natural scale to capture the effects of education. The omission of education from the regression models requires either a compelling argument for why it should not be included, or strong empirical evidence that education does not explain the social behaviors of interest to any meaningful extent.

H&M supply four reasons for why "the role of education versus IQ as calculated by a regression equation is tricky to interpret" (p. 124). They assert that

1. education is at least partly caused by intelligence,
2. effects of education are likely to be discontinuous, that is high school or college graduation might be meaningful but not years of education,
3. *multicollinearity* (that is the degree to which independent variables are correlated) might lead to unstable and misleading regression estimates, and
4. the effects of education and intelligence are likely to be complex and require more complicated modeling.

Assertions 3 and 4 were treated in some detail earlier in the sections on the independence of independent variables and logistic regression coefficients. We saw that the same arguments hold when we consider the correlation and complex effects of IQ and SES-- either the role of SES versus IQ is also "tricky to interpret," which is probably the case, or these two arguments against the inclusion of education should not hold. H&M simply cannot have it both ways. Assertion 2 is nothing more than a technicality easily handled by including education in the regressions as a categorical variable with three levels: less than high school, high school, college or higher education. Moreover, by comparing results from using years of education against results from using this trichotomy, one could directly test assertion 2. H&M use this technique successfully to estimate the effects of Cognitive Classes, rather than a continuous IQ score (see p. 587).

Assertion 1 hypothesizes a causal link, whereby IQ determines the number of years of education completed . In Appendix 3, H&M present an alternative - they entertain the hypothesis that IQ gains are caused by years of education, and note that "it might be reasonable to think about IQ gains for six additional years of education when comparing subjects who had no schooling versus those who reached sixth grade, or even comparing those who dropped out in sixth grade and those who remained through high school" (p. 591). The cause and effect relationship between IQ and education is admittedly complex and open to competing interpretations, but we are not given compelling argument or empirical evidence to support the dismissal of education and the inclusion of IQ in the regressions because of these complex relationships. We can just as validly argue for the inclusion of education and the dismissal of IQ from the regressions. One last point: if years of education as an independent variable competing with IQ for explanatory power, causes H&M so much concern, shouldn't they also worry about the fact that years of education constitute half (and sometime more) of the parental SES index? Surely, assertions 2-4, if valid, pose similar problems for the interpretation of the role of IQ versus SES.

What about empirical evidence? H&M's solution to the problems they raise is to run the IQ versus SES regressions separately for those who completed 12 years of education--the high school sample--and those who completed 16 years of education--the college sample. For college graduates, no matter what their IQ is, the risk of poverty is practically zero. (H&M do not show regression results for the college sample in Appendix 4-- these are meaningless when only six of these subjects were in poverty, but they still plot the regression lines in p. 136.) For the high school sample, H&M notice similar patterns for IQ and SES as were previously observed for the entire sub-sample. IQ has a strong effect regardless of SES; SES has much weaker effect. They conclude: "Cognitive ability still has a major effect on poverty even within groups with identical education" (p.137). These analyses, however, do not answer the important question about education: What happens to the effect of IQ after "accounting for" years of education? Restricting the analysis to a homogenous sub-group in terms of educational attainment provides partial and highly misleading information about this question. When "years of education" is entered into the regression, one finds that it is a highly significant predictor of the likelihood of poverty (a regression coefficient of -.40), independent of IQ, and, even more importantly, the coefficient for IQ drops from -.84 to -.63. However, an even better solution exists.

Responding to criticisms about the SES scale, Murray poses a challenge:

> "Create some other scales and use some other method of combining them.... As scholars are supposed to do, Herrnstein and I checked out these and many other possibilities - the results reported in *The Bell Curve* were triangulated in numbing detail over the years we worked on the book - and we knew that the critics who bothered to retrace our steps would discover: that there is no way to construct a measure of socioeconomic background using the accepted constituent variables that makes much difference in the independent role of IQ" (1995, p. 29).

The following exercise does the obvious. Given the strong correlation between subjects' years of education and parents' SES, and considering that doubtless the most direct way in which parental socioeconomic status can be translated into meaningful advantages for their children is to enable them to get more (and better) education, why not combine these two variables to achieve a better measure of SES? The gains are clear: we increase the SES index reliability, we avoid having three highly correlated variables in the same regression, we update the scale to capture directly at least part of the subjects' realized potential in socioeconomic status. At the same time we resolve some problems of the special case of education. This is achieved simply by averaging the original SES scale with a standardized variable of the subjects' years of education. Table 1 presents the results of the regression of poverty on IQ and the revised SES index.

## Table 1
## Logistic Regression Results Using Revised SES
## (cf. *The Bell Curve*, p. 596)

| | Estimate | Std. Error | t value |
|---|---|---|---|
| **(Intercept)** | -2.695789 | 0.078846 | -34.1905 |
| **IQ** | -0.652195 | 0.106231 | -6.1394 |
| **Revised SES** | -0.622218 | 0.122195 | -5.0920 |
| **Age** | -0.036356 | 0.072727 | -0.4999 |

We can now examine how these new results translate to the plots of IQ versus SES in the roles they play in determining whether young white adults are below the poverty line.

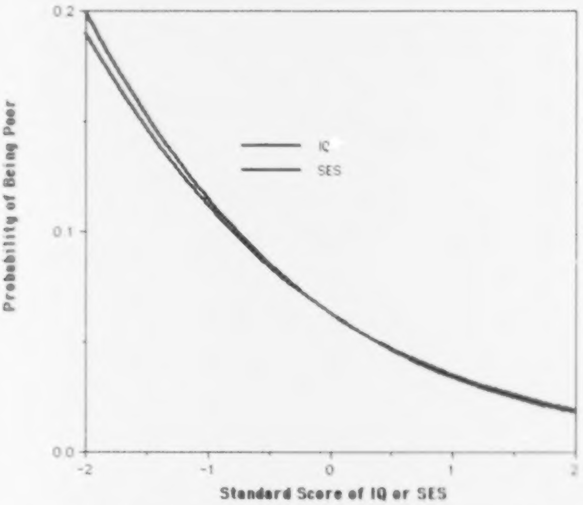

Figure 2. Probability of Being in Poverty as a Function of IQ or SES

This simple and straight-forward improvement of the SES scale - adding the subject's own years of education - brings the relative weights of IQ and SES in predicting poverty to a perfect tie. Dominance of IQ? Hardly. A crucial role for SES? Definitely. Especially if we recall, as H&M themselves acknowledge, that "[SES] has a significant effect on IQ, independent of the genetic endowment of the parent" (p. 589). Moreover, this finding has devastating consequences for any argument about the dominance of the inherited portion of intelligence, 60 percent is the estimate favored by H&M (see p. 105), over environmental factors in determining the odds of being poor. Remember the question we started with? "If you have to choose, is it better *to be born* smart or rich?" (p.127; italics added). The answer is left to the reader.

Does the revised SES and IQ model should be considered adequate for making sound inferences about the relationships among socioeconomic background, education, intelligence, and social behavior? Certainly not. In reality, the social scientist faces an almost insurmountable task when trying to disentangle and bound causes and effects that present themselves only indirectly as a complex pattern of things that go together. Rich families provide better home environment and better education for their children, children with better home environment and better education do better on IQ tests, students who do better on IQ tests are more likely to complete more years of education, they are also more likely to come from families who are better off and less likely to end up poor, and so on and so on. The biggest fallacy behind *The Bell Curve* statistical analyses in Part II of the book is summarized by H&M in a single statement: "Regression analysis tells you how much each cause actually affects the result, *taking the role of all the other hypothesized causes into account*" (p. 122; italics in original). If nothing more, this commentary should provide a demonstration of the dangers of blindly replacing hard thinking about a problem with an analytical formality, sophisticated as it may be.

## Conclusion

In a response to *The Bell Curve*'s critics, Charles Murray repairs to scientific middle-of-the-road and claims: "The statistical method we use throughout is the basic technique for discussing causation in nonexperimental situations: regression analyses, usually with only three independent variables. We interpret the results according to accepted practice" (1995, p. 27). Still, it appears that the analyses of relationships among IQ, SES, education, and poverty suffer in *The Bell Curve* from H&M's quest for simple answers. H&M prefer to ignore important details of their analyses, treat their models and estimated parameters as if they were accurate and complete descriptions of social reality, and pretend that statistical methods can miraculously unravel or unequivocally differentiate among causes that are inherently confounded .

The inconsistencies and selectiveness in arguments and analysis choices documented in the current commentary lead one to wonder whether H&M were not investing too much of their own IQs to make the case for the dominance of intelligence stronger than it really is? Otherwise, many of their conclusions, especially the ones they push about the proper policy response to ethnic and racial differences, lose critically in weight and can hardly be sustained by less extravagant demonstrations of the over-arching importance of IQ in the allocation of opportunities in current American society.

It is only appropriate to end by rephrasing Murray's words: "The unfounded criticisms of the statistics in *The Bell Curve* ... will merely cause embarrassment among a few who both understand the issues and have the decency to be embarrassed" (1995, p. 28). It is my hope that the founded criticisms of the statistics in *The Bell Curve*, will not merely cause embarrassment to its author, but will encourage those "who both understand the issues and have the decency" to set the record straight.

## References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Cronbach, L.J. (1975). Five decades of public controversy over mental testing. *American Psychologist,* 30, 1-14.

Devlin, B., Fienberg, S.E., Resnick, D., & Roeder, K. (1995). Galton redux: Intelligence, race and society, *Journal of the American Statistical Association,* 90, pp. 1483-1488.

Efron, B, & Tibshirani, R.J. (1993). *An introduction to the bootstrap.* New York: Chapman and Hall.

Gustafsson, J-E., & Muthen, B.O. (1994). The nature of the general factor in hierarchical models of the structure of cognitive abilities: Alternative model tested on data from regular and experimental military enlistment tests. Gothenberg University, Sweden: Unpublished manuscript.

Herrnstein, R.J., & Murray, C. (1994). *The Bell Curve: Intelligence and class structure in american life.* New York: The Free Press.

Jensen, A.R. (1981). *Straight talk about mental tests.* New York: Free Press.

McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models.* New York: Chapman and Hall.

Mosteller, F., & Tukey, J.W. (1977). *Data analysis and regression.* Reading, MA: Addison-Wesley.

Murray, C. (May, 1995). "*The Bell Curve*" and its critics. *Commentary*, 23-30.

Neter, J., Wasserman, W., Kutner, M.H. (1990). *Applied linear statistical models.* Homewood, IL: Irwin.

## About the Author

**Haggai Kupermintz**

School of Education
Stanford University

Haggai Kupermintz is a doctoral candidate in Psychological Studies in Education, School of Education, Stanford University, Stanford, CA 94305. His specializations are educational measurement and statistics.

haggaik@stanford.edu

Home Page

---

**Copyright 1996 by the *Education Policy Analysis Archives***

*EPAA* can be accessed either by visiting one of its several archived forms or by subscribing to the LISTSERV known as EPAA at LISTSERV@asu.edu. (To subscribe, send an email letter to LISTSERV@asu.edu whose sole contents are SUB EPAA your-name.) As articles are published by the *Archives*, they are sent immediately to the EPAA subscribers and simultaneously archived in three forms. Articles are archived on *EPAA* as individual files under the name of the author and the Volume and article number. For example, the article by Stephen Kemmis in Volume 1, Number 1 of the *Archives* can be retrieved by sending an e-mail letter to LISTSERV@asu.edu and making the single line in the letter read GET KEMMIS V1N1 F=MAIL. For a table of contents of the entire ARCHIVES, send the following e-mail message to LISTSERV@asu.edu: INDEX EPAA F=MAIL, that is, send an e-mail letter and make its single line read INDEX EPAA F=MAIL.

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

To receive a publication guide for submitting articles, see the *EPAA* World Wide Web site or send an e-mail letter to LISTSERV@asu.edu and include the single line GET EPAA PUBGUIDE F=MAIL. It will be sent to you by return e-mail. General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, Glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692)

### Editorial Board

Greg Camilli
*camilli@pisces.rutgers.edu*

John Covaleskie
*jcovales@nmu.edu*

Andrew Coulson
*andrewco@ix.netcom.com*

Alan Davis
*adavis@castle.cudenver.edu*

Sherman Dorn
*sfxj9x@scfn.thpl.lib.fl.us*

Mark E. Fetler
*mfetler@ctc.ca.gov*

Thomas F. Green
*tfgreen@mailbox.syr.edu*

Alison I. Griffith
*agriffith@edu.yorku.ca*

Arlen Gullickson
*gullickson@gw.wmich.edu*

Ernest R. House
*ernie.house@colorado.edu*

Aimee Howley
*ess016@marshall.wvnet.edu*

Craig B. Howley
*u56e3@wvnvm.bitnet*

William Hunter
*hunter@acs.ucalgary.ca*

Richard M. Jaeger
*rmjaeger@iris.uncg.edu*

Benjamin Levin
*levin@ccu.umanitoba.ca*

Thomas Mauhs-Pugh
*thomas.mauhs-pugh@dartmouth.edu*

Dewayne Matthews
*dm@wiche.edu*

Mary P. McKeown
*iadmpm@asuvm.inre.asu.edu*

Les McLean
*lmclean@oise.on.ca*

Susan Bobbitt Nolen
*sunolen@u.washington.edu*

Anne L. Pemberton
*apembert@pen.k12.va.us*

Hugh G. Petrie
*prohugh@ubvms.cc.buffalo.edu*

Richard C. Richardson
*richard.richardson@asu.edu*

Anthony G. Rud Jr.
*rud@purdue.edu*

Dennis Sayers
*dmsayers@ucdavis.edu*

Jay Scribner
*jayscrib@tenet.edu*

Robert Stonehill
*rstonehi@inet.ed.gov*

Robert T. Stout
*stout@asu.edu*